THE CENTURY PSYCHOLOGY SERIES

RICHARD M. ELLIOTT, *Editor*
KENNETH MACCORQUODALE, *Assistant Editor*

Experimental Psychology

Experimental Psychology

AN INTRODUCTION.

BENTON J. UNDERWOOD
Associate Professor of Psychology
Northwestern University

THE TIMES OF INDIA PRESS
BOMBAY

PUBLISHED IN INDIA BY THE TIMES OF INDIA PRESS, BOMBAY,
BY ARRANGEMENT WITH APPLETON-CENTURY-CROFTS,
DIVISION OF MEREDITH PUBLISHING COMPANY, NEW YORK, N.Y., U.S.A.

This book has been published with the assistance of the Joint Indian-American Standard Works Programme.

Rs. 9.00

PRINTED AND PUBLISHED FOR THE PROPRIETORS BENNETT, COLEMAN & CO.,
LTD., BY PYARELAL SAH AT THE TIMES OF INDIA PRESS, BOMBAY, INDIA.

Preface

In 1946 I was faced with the problem of teaching a course in undergraduate experimental psychology with no available text seeming suitable for background reading. Therefore, I started to bring together certain materials to be mimeographed and issued to the students as a substitute for a text. These materials were revised and expanded from quarter to quarter until the present final draft was written.

I am of the opinion that subject matter (content) and methodology should not be divorced in an experimental course; one without the other is deadening. However, since I have done no acceptable research on these problems, my opinion may be in error. And, although I set up the objective of continuous unification of method and content, I must add quickly that I was unable to attain the objective completely. At some point in the writing it became apparent that to keep the text to a reasonable size some omissions would be necessary. The major omissions, as I see them, should be pointed out. (1) Factual data on the sensory (discriminal) processes. I found that I could not handle adequately the factual relationships which have stemmed from the great flow of research in this area and still give a comprehensible account of the methods. I chose methods. (2) Physiology of sense organs and nervous system. Because this material is not unique to psychology it was excluded. (3) Details of apparatus construction. I have eliminated nearly all engineering from the book; references to a piece of apparatus are in terms of how it is used, not how it is made.

In mimeographed form the text has been used in a two-quarter course for which Elementary Statistics was a prerequisite, and in a year course where statistics was taught as an integral part of the methodology. I believe that the factual subject matter can be

comprehended readily without a statistical knowledge, but a full appreciation of experimental design problems requires some statistical thinking.

The experimental illustrations used throughout are, for the most part, recently published. In most areas I believe the temper of contemporary research is accurately reflected. In some instances I have attempted to guess the future trends of experimentation. It will also be noted that I have reported some experiments which in my opinion were poorly conceived, executed, or interpreted. I think it is clear that I have done this for pedagogical reasons and with no thought of malice.

The entire final draft was read critically by R. M. Elliott, Editor of the Series, and by Claude E. Buxton and Kenneth MacCorquodale. Parts of the final draft were read by R. W. Kleemeier and C. P. Duncan, who were also kind enough to use the mimeographed edition in their classes and gather student reaction as well as make suggestions of their own for revision. To realize fully my indebtedness to all these men I need only glance over the many pencilled notes which they made on the drafts. Louise Underwood has contributed toward the clarifying of expression and has helped with the proofs. Marion Cisar checked the references and was able also to type and "unsplit" infinitives simultaneously.

Permission has been granted by many publishers and editors to reproduce figures or passages, and I gratefully acknowledge these permissions: The American Psychological Association for its several journals; Prof. R. S. Woodworth for *Archives of Psychology;* Prof. K. M. Dallenbach for *American Journal of Psychology;* Dr. Carl Murchison for *Journal of Psychology, Journal of General Psychology,* and *Journal of Genetic Psychology;* the editors or publishers of *Journal of Educational Psychology; Journal of Comparative Psychology; Psychiatry;* and *British Journal of Psychology,* and the following publishing firms: University of Chicago Press, University of Iowa Press, Duke University Press, Columbia University Press, Yale University Press, National Society for Study of Education, Appleton-Century-Crofts, Inc., Henry Holt & Co., Rinehart & Co., Longmans, Green & Co., Ronald Press Company, and Harper & Bros. The original source is given in each case.

Evanston, Illinois B. J. U.

Contents

Experimental Psychology

CHAPTER I

Introduction

THE BASIC CONCEPTS OF EXPERIMENTAL PSYCHOLOGY

Human responses are the basic data of study by psychologists. People have dreams, they get married, they learn the alphabet, they bet on horse races. Why do these responses occur? "*Why* do people dream?" "*Why* do people bet on horse races?" In asking such questions we immediately imply that all these events have a cause and, if we ask the questions sincerely, we presuppose a discoverable cause. We imply that there must be certain conditions which lead to dreaming and to betting on horses. The events which the psychologist attempts to relate or discover are thus evident: *psychologists attempt to determine the relationships between responses of the organism and the events which prompt those responses.*

All sciences attempt to establish relationships between events. It is a part of the task of the scientist to find these relationships and state them in a fashion that will be agreed to by other scientists. That we have different names, such as physics, chemistry, botany, astronomy, psychology, applied to scientific endeavors stems largely from the fact that there are different classes of events within which one may seek to establish relationships. Thus, the physicist is concerned largely with certain events among inanimate objects; the botanist with events relating to the growth and structure of plant life; the zoölogist with growth and structure of animals, and so forth. Different sciences represent divisions of labor; their basic task, however, remains the same.

Stimulus and response. A response is an observable change in behavior, this change being an activity effected by muscles or glands. An event which elicits a response has come to be known

as a stimulus. No adequate general definition of a stimulus can be given without reference to the occurrence of a response. *External stimuli* are originally energy changes in the environment, but not all such changes are stimuli. Only those energy changes which produce responses can be said to be stimuli.

We are prone to think of stimuli as events originating outside the organism: a loud noise causes a startle; a red light produces a brake-depressing response. While it is true that many of our responses are initiated by external energy changes to which our receptors (sense organs) are sensitive, it must be recognized also that many stimuli are *internal* in origin. Hunger pangs are just as truly response-producing as are external stimuli. We must not, then, limit the conception of stimuli to external changes. Internal stimuli not only influence behavior but some may also be indirectly controlled and varied by manipulating the environment outside the organism. Since we are to study methods by which relationships between responses and conditions prompting those responses are determined, techniques of controlling and manipulating both external and internal stimuli must come under our surveillance.

Experimental framework. In diagrammatic form (Fig. 1) we may suggest the experimental framework of psychology. In the experimental situation it is recognized that there are a multitude of stimuli (S_1, S_2, . . . S_n) impinging upon the organism. It is also recognized that there are identifiable and unidentifiable conditions within the organism which may influence the response or responses we are interested in measuring. *The basic experimental procedure is that we hold all stimuli constant except one* and this one we manipulate or vary as our demands dictate. If we can show that the response changes in some systematic fashion as the stimulus changes, we have established one causal condition of a psychological event.

Figure 1 indicates that there may be *many* responses (R_1, R_2, . . . R_n) made by the organism upon which we are experimenting. The experimenter, then, arbitrarily measures a response (or responses) in which he is particularly interested and which he thinks is some function of the stimulus he is manipulating, i.e., is dependent upon the stimulus. To be scientifically useful, all that is required is that the response be in some manner measur-

able. It is possible, of course, as some have contended (130),* that many extremely significant responses are overlooked or ignored in many experiments. Significant responses would be those which, if measured, would give us greater understanding of the phenomenon being investigated.

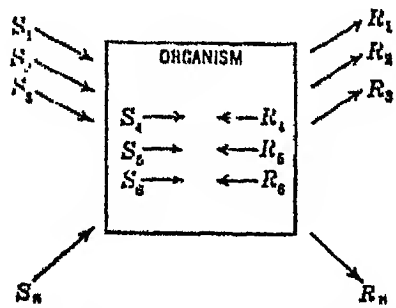


FIG. 1. The experimental framework of psychology. $S_1, S_2, \ldots S_n$ indicate stimulus events of an unknown number, both in the environment and in the organism. $R_1, R_2, \ldots R_n$ indicate responses of the organism to the stimuli. Responses "turned back in" the organism (R_4, R_5, R_6) take note of the fact that responses may in turn become stimuli influencing later responses. Experimental psychology at the empirical level is concerned with the discovery of the laws which hold between the stimulus events and the response events. If this diagram seems lifeless, consider the square a torso and add legs, arms, and a head.

It was indicated above that in an experiment all stimuli except one must be held constant. One of the fundamental parts of training in experimental technique is to be able to evaluate an experimental design—a detailed outline of a proposed experiment—from the standpoint of how well all stimulus conditions will be controlled, and thus how precisely a response can be said to be caused by a specific stimulus event. These stimulus conditions which are known to have an influence on the responses of the organism are called *stimulus variables.* Stimulus conditions whose influence on a given phenomenon is not known, we shall call *potential stimulus variables.* If all known or potential stimulus variables are not controlled—held constant—at the time another stimulus is being experimentally varied, we confuse the relationship which the experiment is attempting to establish. You will find that only

* Numbers occurring in parenthesis refer to the source of the citation as listed in alphabetical order at the end of the book.

a careful and critical scrutiny of past experiments, some carefully and some poorly designed, will allow one to develop a "feel" for sound procedure. Exercises to enhance this training should be a part of your laboratory experiences.

ILLUSTRATIONS OF THE BASIC CONCEPTS

Let us consider some simple experimental designs from the standpoint of their adequacy of control of stimulus conditions. At the same time we will get illustrations of the concepts which have been introduced. Since experimental procedure is basically the same for most sciences, the first hypothetical experiment is taken from the field of horticulture.

A florist specializes in orchids. For many years he has raised orchids of several different varieties. It has occurred to him that perhaps all of his strains are not equally productive. For example, he has three strains of orchids all of which produce yellow blossoms. The question, and a very practical one, which he puts to himself is: "Do all of these three strains produce an equal number of blossoms of the same size?" If there are differences it might be worth while financially to find out which seed produces most efficiently. An experiment is devised to help provide the answer.

Now, there are several "stimulus" variables which the florist has to control. For example, we may name the amount of moisture each seed gets from planting time to maturity. Since moisture influences plant growth, the amount of this variable would have to be the same for all three strains. The fertility or chemical composition of the soil or solutions would have to be the same for all plots in which the seeds are planted. A soil analysis might be made in order to insure this. There are other variables, such as the amount of sunlight, the temperature, the slope of the ground, and so forth, which may be known or suspected to influence the growth differentially. They would, therefore, have to be the same for all seeds. Obviously not just one seed would be used but rather a good many, so that there would be a good representation of each strain. In summary of his procedure, then, the florist knows that in order to get a clear-cut answer to his question, he has to keep all factors (stimulus variables) constant except one—

the type of seed. This variable, the type of seed, is his experimental or manipulated variable, or as it is usually called in psychological experiments, the *independent variable. The florist's procedure gives all seeds an equal opportunity to produce orchids.* This is controlled experimentation.

To carry the experiment through to a satisfactory conclusion the florist would maintain constant conditions throughout the growth period (which is actually several years for orchids), and then attempt at the appropriate time to measure which strain of seed made the "best response." If his judgment of these responses is unbiased, he will be able to provide himself with at least a restricted answer to his original question. It is necessarily a restricted answer—the generalization he can make is sharply limited—because he does not know (without further experimentation) what seed would produce the finest orchids under conditions other than those which prevail in this experiment. It is possible, for example, that at a different temperature level another seed would produce more efficiently. Nevertheless, he can state a reliable finding—a relationship between his seeds on the one hand and the products of those seeds on the other. This relationship is not subject to doubt if he carries his experimental procedure through in a carefully controlled manner.

Now let us suppose that the florist did not control the moisture of the soil. Without thinking, perhaps, he may have watered one plot of ground more than the others. If this were the case, we would challenge the purported relationship which he thought his experiment demonstrated. We might insist: "But you don't know whether the differences in your orchids were a product of the differences in the seeds as such or of the differences in the moisture provided to those seeds." In a true scientific sense he has no answer to such a comment; his experiment has "proved" nothing except the need for more careful procedure.

It may be stated categorically that errors in the design of psychological experiments are relatively easy to make. Psychological experiments, as compared with the florist's experiment, are really considerably more difficult to evaluate because there are so many stimulus conditions—stimulus variables—that it is often difficult to know whether or not we have them all controlled. It cannot be stressed too strongly that the beginning experimentalist must

develop a critical eye for experimental defects if he expects to carry on sound research or evaluate other research. Like all learned behavior, this must be practiced constantly.

Now consider a hypothetical psychological experiment which parallels the one performed by the florist. An advertising agency, suspecting that there must be differences in the efficiency of various types of printing, decides to run an experiment to discover which of three types of print is the most quickly read by an adult. This problem is a real one to the agency, since the outcome of advertising depends in part upon how quickly the reader gets the message which the advertisement intends to convey. If one form of print is the most efficient, it could be used most frequently in the layouts.

The independent variable in this advertising experiment is the type of print. This stimulus condition will be varied in three ways by using three different print types. The response variable, the *dependent variable,* will be the speed of comprehension of each of the three print types. What are the stimulus variables which need to be considered and controlled? Or, paralleling the florist's experiment, what do we have to control that is similar to the moisture content, the fertility of the soil, and so forth? Actually, the number of these conditions which would have to be controlled is unknown. Many of them would be known, however. We should think of this problem by asking ourselves the question: "What factors might influence the speed of comprehending the print types in addition to the print types as such?" Such factors fall into two general classes: manipulable stimulus variables and subject variables.

By *manipulable stimulus variables* we mean characteristics of the experimental environment (including the task itself) which the experimenter can vary and control. In this experiment, illustrations of manipulable variables would be such things as the words or sentences used to exhibit the print type; the length of time the material was shown to the subject; the length of the experimental period; time between successive presentations; the amount of light in the room.

By *subject variables* we mean characteristics of the subject which might influence the response. These are almost unlimited in number, examples being chronological age, mental age, sex,

amount of education, visual acuity, and home background. These are all at least potential variables, since it is reasonable to suspect that they might influence the speed with which the material is comprehended.

To execute the experiment we must be sure that all of the manipulable variables, known and potential, are held constant while perceptual speed for each print type is measured. If we use a different group of subjects to measure the perceptual speed for each print type, we must be certain that the groups do not differ significantly on any known pertinent subject variable. We allow only one factor–print type–to vary.*

We shall not attempt at this time to detail one of various methods by which such an experiment could be executed. A careful study of the methods, and reasoning behind them, by which this and many other different kinds of psychological experiments can be performed is one of the major purposes of the entire book. We will probably find it less overwhelming if the methods and their rationale are brought before us in small, but systematic and persistent doses.

DIMENSIONS

We need now to enlarge on the concepts of stimulus and response variables and to add to our basic ideas concerning the framework of experimental psychology.

When any given phenomenon can be demonstrated reliably (consistently) to vary in amount with respect to some specific characteristic, we have a *dimension*. The characteristic may be any one of many things. In the physical world, such characteristics as length, weight, and temperature are common dimensions used in describing natural phenomena. Clearly, objects can be reported reliably as varying in length, weight, or temperature. Does human behavior vary in certain characteristics and can

* Modern statistical discoveries have made it possible to manipulate several variables in a single experiment and still assess the influence of each. Such experimental designs actually consist of several experiments in one, in which each conforms to the basic rule of manipulating a single variable. The statistical techniques required to evaluate the complex designs are typically taught only in advanced courses. Our emphasis, therefore, will be upon the simple case of manipulating only one variable while holding all others constant.

these variations be reported reliably? Certainly, for do we not have variations which can be measured reliably in speed of learning, intelligence, visual acuity, and so forth? Indeed, any word used to describe human behavior indicates a characteristic which might be dimensionalized.

The difficulty which lies in the way of dimensionalizing many behavior characteristics is that reliable measurements of them cannot be made, i.e., the amounts of the characteristic present in individual cases cannot be agreed upon. Take, for example, sense of humor, a characteristic of behavior we often judge in others. We would all agree that different amounts or degrees of this characteristic exist, and yet if we attempted to dimensionalize this characteristic we would probably run into a great many difficulties. We would probably find that we do not agree on what constitutes a good sense of humor or a poor sense of humor and as a consequence, my estimation of X's sense of humor might vary markedly from yours. We are forced to use our judgments to "measure" this characteristic because we have no instruments which will do the job. In this case our judgment as the measuring instrument is quite unreliable. We would all be fairly consistent in our measurement of the height of an individual but quite inconsistent in our "measurement" of his sense of humor. When a characteristic cannot be dimensionalized in the manner we have described, its usefulness as a scientific concept is sharply restricted.

When a characteristic of behavior can be dimensionalized, we then speak of it as a *response dimension*. If the characteristic or characteristics of the stimulus which produces variation in response can be dimensionalized, we may then speak of a *stimulus dimension*. If a pure tone is made to vary in frequency (cycles per second), the subject usually reports that he hears a concomitant variation in pitch. Cycles per second is thus the stimulus dimension, and perception of pitch changes correlated with it make up the response dimension. Or, rate of learning varies as a function of the degree of meaningfulness of the material. Meaningfulness is the stimulus dimension; rate of learning the response dimension.

It now becomes apparent that stimulus dimensions are also stimulus variables as we have used the latter term previously. Stimulus variables are related to variations in response; so are

stimulus dimensions. What is the difference? The difference as used here is that stimulus variables include stimulus conditions which are not dimensionalized. A non-dimensionalized stimulus variable produces changes in the response but we are not able to relate the stimulus changes to the response changes. In the hypothetical experiment on print types we have stimulus variation—three different kinds of print—but the characteristics by which the print types varied were not dimensionalized. Perhaps these print types could have been dimensionalized since they would vary in height, width, and stroke width. As we discussed the experiment, however, we merely had three distinct print types. Our results would be sound enough, but without dimensionalizing the characteristics, we could not describe the relationship between any given characteristic of the stimuli and the responses evoked thereby. When stimulus conditions are grossly judged to be different and when the characteristics in which they differ are not dimensionalized, indeed, often not specified, the differences are sometimes said to be qualitative. An increase in the number of dimensionalized stimulus variables and a decrease in the number of qualitative variables is an indication of scientific advancement.

We are now ready to refine our conception of the experimental framework of psychology. We have seen that all stimulus dimensions are stimulus variables, but that all stimulus variables are not dimensionalized. All response variables, the dependent variables, must be dimensionalized to be scientifically useful. To incorporate the concept of dimensionalized stimulus variables, we modify Fig. 1 as shown in Fig. 2. We have depicted a single stimulus characteristic to indicate that all stimulus conditions which are dimensionalized vary from high to low in terms of amount of that characteristic. A part of the experimentalist's work is to determine the relationships between variations of stimulus dimensions and the corresponding changes in the response dimensions. In actual practice, several rather equally separated points along the stimulus dimension are used as the specific stimulus conditions. Plotting these amounts against the amount of corresponding response change (if any) shows the empirical relationship. It must be understood at this point that Fig. 2 is not intended to convey the impression that there is necessarily a one-to-one relationship between variations in the stimulus and variation in

the response. The work of the experimental psychologist would be much simpler if such relationships did hold.

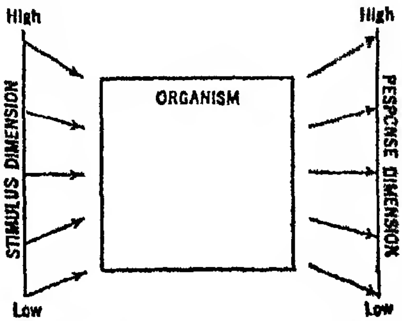


FIG. 2. Refinement of the framework of experimental psychology to include the concepts of stimulus and response dimensions. For each stimulus dimension and for each response associated with it, this representation would apply. The diagram suggests a one to one relationship between the stimulus and response dimension. This is rarely the case, but the figure is drawn this way for greater simplicity.

Physical and psychological stimulus dimension. In the study of many psychological processes the stimulus dimension is a physical scale, being stated in terms of inches, cycles per second, decibels, and so forth. Such stimulus dimensions are called *physical stimulus dimensions*. We have other stimulus dimensions which are peculiarly psychological—*meaningfulness* is an illustration—and these may be called *psychological stimulus dimensions*. Chapters II and III will deal largely with the relationships obtaining between physical stimulus dimensions and response dimensions; Chapter IV will be concerned primarily with relationships between psychological stimulus dimensions and response dimensions for our more simple response categories. Response dimensions, of course, are always to be thought of as psychological dimensions. If one prefers, the word *continuum* may be used synonymously with the word *dimension*.

Questions must arise concerning the assigning of numbers to dimensions. Objects or events of the physical world are almost always described by reference to numbers. The wing span of the plane is 42 feet, 6 inches; the temperature of the room is 68 degrees Fahrenheit; the load of coal weighs 4205 pounds; the speed of

the turbo-jet plane is 462 miles-per-hour. Characteristics which are dimensionalized are assumed to vary in infinitely small amounts, though the measuring unit may vary markedly in magnitude. The unit of length measure may be as small as 1/10,000 of an inch for certain stituations, whereas for others a measurement in miles may be used. Whether or not numbers are assigned to a dimension depends upon several considerations; however, the assignment of numbers is not necessary, it is only desirable. The penchant which scientists have for metricizing or numberizing dimensions of description is due to the fact that the finer the quantification the better the description of the phenomenon in question; hence, a more precise prediction of it can be made for a given set of conditions. *We may, however, have useful dimensions for describing phenomena and use only words to identify points along the dimension.* We could, if necessary, have a dimension of length without using numerical units. We could agree that the Empire State Building is taller than the Brooklyn Bridge, that the Brooklyn Bridge is taller than an oak tree. In summary, then, it may be said that we assign either numbers or words to successive points along a dimension to meet the requirements of a dimension.

THE ORIGIN OF EXPERIMENTAL PROBLEMS

I-wonder-what-would-happen type. Many students, first faced with the task of *finding* a problem to "do an experiment on" are at a loss. Many candidates for the Master's Degree often grope for a problem which they think should be attacked experimentally. Yet, once they get beyond this stage of having done their first major experiment, they are sometimes bothered by not having enough time to do all the experiments they have outlined. Many beginning students have often asked: "How do people think of all these experiments?"

There is really nothing magical in the processes by which a problem comes under experimental investigation. So often in the past, ironically enough, psychological problems which have been attacked experimentally have been so divorced from practical situations that the experimenter was unlikely to develop a "felt need" as a result of his personal contact with the everyday environment. Yet, any question, any problem, or any "felt need"

for an answer to a question about human behavior is potentially an experimental problem. In some cases, of course, it is not possible to devise an experiment that will answer the question.

It seems true, especially in the early development of a science, that many of the experiments are the result of the I-wonder-what-would-happen-if-I-did-this type of question. One might ask: "I wonder which would be most efficient, having my students study by reading each paragraph three times before going on to the next, or having them read the entire assignment over once, then again, then the third time?" The industrial relations director might ask: "I wonder what would happen if we paid our employees once a month instead of once a week?" The comparative psychologist might ask: "I wonder how it might affect the speed of learning if I fed the rats only after every other trip through the maze?" You might ask: "I wonder which would be best for me, to study 15 minutes, rest 15, and study 15 more, or study the full 45 minutes?"

In more technical language the I-wonder-what-would-happen type of question is seeking to know whether or not the factor in question is an effective stimulus variable, i.e., will it influence the response?

I'll-bet-this-would-happen type. As more facts are accumulated within a science the I-wonder-what-would-happen type of problem gives way to what we may call the I'll-bet-this-would-happen-if-I-did-this type. In other words, the experimenter is making a shrewd guess, or if we dress the language up a bit, he is formulating an hypothesis as to the effect of the variation of a certain stimulus variable. His experiment will test the hypothesis. The experimenter already has certain facts and from the implications of these facts he decides, or deduces, that if he performed another experiment of a certain nature the result would be such and such. This procedure, to most scientists, is a fascinating game. Here also, the experimentalist is not so interested in discovering new variables but rather in determining the exact influence of various amounts of a condition known already to be a significant variable.

The I'll-bet-this-would-happen type of problem is usually generated, not from facts alone, but from theories which the scientist formulates to account for those facts. Theories attempt to explain obtained facts and if the theory is a "good" theory, it will go

further and suggest new relationships which should be found. Finding these new relationships is thus the function of experimentation. Empirically discovered relationships, then, do not in themselves explain the basic mechanisms which are back of those relationships. This is the function of theory. Let us consider an illustration.

One of the best-known empirical relationships in psychology is that with the passage of time, forgetting takes place (there are a few exceptions). Time, in this relationship, is the independent variable and the amount forgotten the dependent variable. Are we then to say that forgetting is caused by the passage of time? We may, but it is an exceedingly low level of explanation. Independent variables are not customarily used as explanatory concepts. What we would need in the way of explanation, hence in the way of theory, is some process which changes in a systematic fashion *with the passage of time* and which could bring about forgetting. At present the single most generally favored theoretical concept to explain forgetting is that of interference. During the passage of time we learn other things which interfere with the material already learned. For illustrative purposes, assume that the only empirical relationship we have at hand is that between the passage of time and the amount of forgetting. Someone proposes the interference theory. Now note that such a theory immediately suggests other experiments to determine other empirical relationships. Of course someone has to do the suggesting, but the point is that the theory demands that other relationships must hold if the theory is sound. For example, if the theory is tenable it should be found that the forgetting of material is faster the greater the amount of subsequent material learned. If adequate experimental tests of this specific proposition do not confirm it, we would be forced to throw out the general theory of interference or else modify it drastically. A good theory, then, will not only explain the facts at hand but will also suggest new relationships. The most useful theories are those that clearly point toward empirical relationships which can be determined experimentally.

The more facts we have the more generalized or expansive our theory becomes and the greater the number of experimental problems which are in turn suggested by the theory. It seems likely

that few scientists have to do more than one or two experiments before countless others are suggested to them by the facts they have gathered and by the theory they have formulated to account for those facts.

Good scientific theory develops the I'll-bet-this-would-happen stage of scientific advancement. It is reasonable to presume that experiments suggested by this kind of thinking generally indicate that a science has passed out of the kindergarten stage, for in order to formulate reasonable theory the scientist must have a certain number of well-established facts at hand. These facts are the product of prior experimentation. For most scientists, experimentation is a sterile occupation unless some theory is involved. Theory is the unifying element of science; theory directs experimentation; theory provides broad generalizations for organizing masses of data; theory makes of science a comprehensible affair which a mere array of facts will not; testing a theory often provides the powerful motive needed to get laborious experimentation completed.

Psychology is getting well launched into the I'll-bet-this-would-happen stage. Most of the experiments to be reported in this book were predicated on some theory, although lack of space will prohibit any extensive consideration of these theories. However, because of the basic nature of learning, we shall devote one chapter to experiments built around theories of learning, thus illustrating the tie between theory and experimentation.

STEPS IN EXPERIMENTATION

Most experimental procedures fall quite naturally into four rather distinct steps: (1) Statement of the problem, (2) Design of the experiment, (3) Performing the experiment, and (4) Interpretation of the data.

Statement of the problem. The problem may be stated in question form but more likely it will take the form of an hypothesis. Thus, a theory is stated and an experimental relationship, deduced from the theory, is proposed. The experiment then attempts to discover the degree of relationship which exists in fact and thus supports or denies the theory. It is obvious that the theory and experimental problem must be related in an unambiguous fashion.

Negative results, i.e., results which do not confirm the hypothesis, do not always discredit a theory. Certain factors may have been discovered during the execution of the experiment which were not anticipated during the planning. In such cases the experiment may be said to be an inadequate test of the hypothesis. Usually, careful planning of the experiment, following careful consideration of the propositions suggested by the theory, will reduce such incidents to a minimum.

Beginning experimentalists sometimes confuse the material being used to determine a relationship with the relationship itself. For example, in one laboratory course we were attempting to determine the relationship between the degree of meaningfulness of the material to be learned and the rate of learning. Nonsense syllables were selected as the learning material. These had been sorted for differences in meaningfulness in an earlier experiment. In the written report of the experiment one student indicated the problem as being: *the learning of nonsense syllables.* To this particular student the learning of nonsense syllables *was* a problem and he had apparently confused his personal activities and difficulties with those of becoming an experimentalist.

Let us take the above problem and see how it might be stated in several forms.

1. He might lose the question as an I-wonder-what-would-happen type. He would state the purpose of the experiment as an attempt to get an answer to the question: *Does meaningfulness influence the rate of learning?*

2. The I-wonder-what-would-happen type of approach may also be stated in more explicit fashion as follows: *To determine the relationship between meaningfulness of moterial and the rate of learning as shown by the learning time of nonsense syllables of different degrees of meaningfulness.* This is an explicit statement of the problem and also indicates the general way it is to be tested. The statement of the problem assumes in effect that there is a relationship, and the purpose of the experiment is to discover exactly what that relationship is.

3. If he stated the problem in hypothesis form he would say: *The problem is to test the hypothesis that the rate of learning is a direct function of the degree of meaningfulness of the material to be learned.*

Which one of the above three forms one uses is not determined on any "right" or "wrong" basis, but rather by the amount of knowledge the experimenter already has concerning the problem, plus any personal likes he has for stating problems in certain ways.

Designing the experiment. The general direction or general form the experiment takes is usually suggested by the question asked or the hypothesis stated, but the precise planning of the details is a different matter. Here is where experimental skill shows itself. All potential stimulus variables must be controlled. Here also a statistical knowledge is necessary so that the data to be gathered can be stated in acceptable quantitative terms. Always in designing an experiment we keep two general questions in mind: "Will this give a satisfactory answer to the question or be an adequate test of the hypothesis?" and "Do I have all potential stimulus variables controlled except one?"

This step in the experimental procedure is the major training ground for the neophyte experimentalist.

Performing the experiment. If the experiment has been designed carefully and if the experimenter is fully aware of the necessity of maintaining constant conditions, this step is of a routine nature. Certain skills may be necessary, especially if the apparatus used is complex. Experiments may fail during this stage simply because the experimenter is unaware of gradual shifts in his procedure. Many of the automatic laboratory devices now in use eliminate much of the human control, but even these are subject to errors that may confound the results even if the experimenter is skilled in the use of the apparatus. Many beginning students of experimental procedure find the constant repetition of conditions (which is necessary in many experiments) exceedingly boring, and indeed it usually is. However, this should never be allowed to affect the attitude toward the procedure or toward the subject. Even casual remarks to the subject, or to an occasional subject, may produce variation in the results. For example, a subject may be learning a list of words under standard laboratory conditions. If by any comment or expression from the experimenter, no matter how subtle, the subject gets an idea that his ability to learn the list of words is a measure of his intelligence, the results may be considerably different. There is no reason why laboratory conditions should change as a function of the experimenter if that

person is fully and constantly aware how absolutely essential it is to maintain rigid conditions.

Interpretation of the data. In most cases interpretation consists of making the statistical computations and evaluating the results of those computations so that one obtains an asnwer to the question prompting the experiment, or a confirmation or negation of the hypothesis being tested. Facility in this area will usually follow automatically the skill developed in carrying out step two. In addition, however, the data may suggest extensions of the original theory or it may suggest new theory. How does one learn to theorize? It is a good guess that we learn this skill in the same manner we learn anything else—by practice.

In writing experimental reports it is customary to follow somewhat the same order as employed in carrying out the experiment. Five sections may be used as follows:

1. Statement of the problem. This section may include background material, such as a summary of related studies and a presentation of the theory with which the experiment is to be concerned.
2. Detailed statement of the procedure used.
3. Presentation of the results. Graphs and tables are commonly used to summarize the basic findings.
4. Interpretation or discussion of the findings with their implications for theory and future research.
5. Conclusions and summary.

PLAN FOR INTEGRATING AREAS OF STUDY

We have two major objectives to accomplish in this book: (1) To obtain a knowledge of the influence of known major variables which have been shown experimentally to influence behavior; (2) To get a critical appreciation of sound experimental design. The first objective points toward factual subject matter, the acquisition of which is mandatory for reaching the second objective, since we can best design experiments if we know what variables need to be controlled.

It has become customary to approach experimental psychology stepwise through different areas of subject matter and this is what we shall do. It has also been customary to find that most students

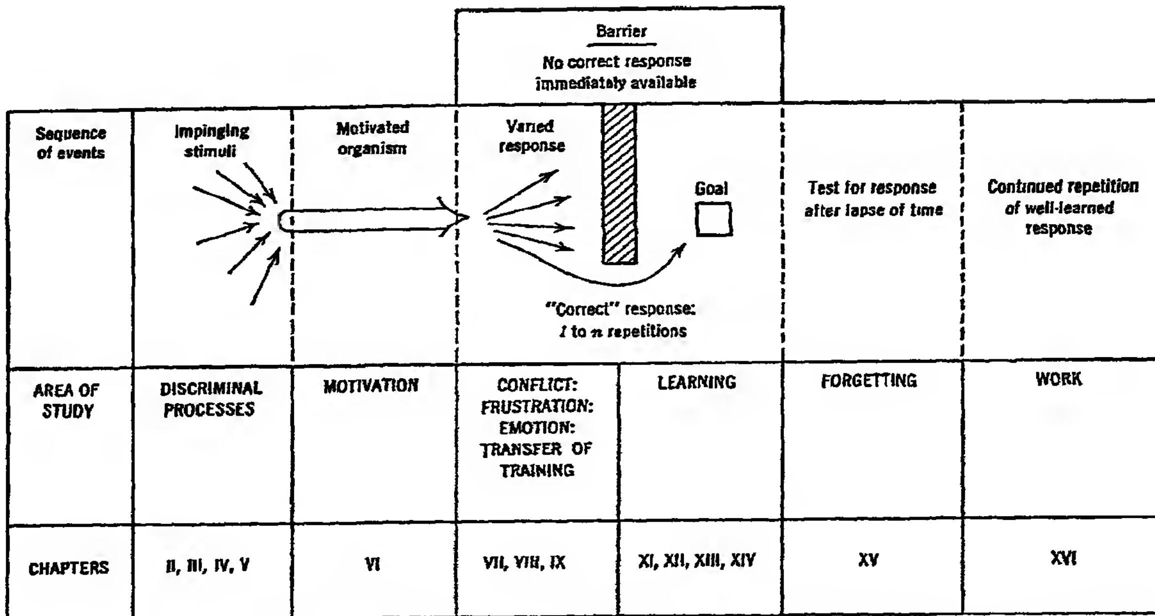


FIG. 3. Areas of study depicted as a sequence of events. The diagram should be circular in part. Thus, in so far as motives are learned, the sequence should point back upon itself. We should recognize that many variables are active in producing each of the individual phenomena occurring at the various points in the sequence.

have difficulty in integrating the areas of experimental psychology into a relatively simple conceptual structure. Of course one could know in advance that no structure is going to be completely satisfactory. For the areas to be covered in this book the author feels that the basic diagram proposed by Dashiell (76) and since used by Melton (254) and Shaffer (348) is fairly adequate. This diagram, considerably elaborated, is shown in Fig. 3.

First, we need to know the methods which have been used to study the organism's sensory and perceptual capacity. We should know the methods which have been used to determine the characteristics of the responses evoked by physical stimuli to which the organism is sensitive. We should know something of the factors which determine simple judgments and sensory discriminations. We shall call all of these by the term, *discriminal processes*.

The next area of study is *motivation*. This is a major variable in all behavior. The arrow in Fig. 3 is intended to indicate that the organism is motivated toward a goal.

The hypothetical organism with which we are dealing in Fig. 3 has no ready response to the stimulus complex (including the motive); a smooth, well-developed behavior pattern will not appear. This situation, of no immediately available response, is represented by the barrier. The behavior which results from even the temporary blocking of goal-directed behavior has many experimental facets. Here we will be concerned with experimental work on *frustration* and *conflict* and associated *emotional states*. When a behavior sequence is blocked the organism usually shows trial behavior, varying its response until the "correct" response is discovered. What the organism does at the barrier point seems to be in large part a function of its past experience. We shall study the particular problem known as *transfer of training*.

Continuing from left to right in Fig. 3, we can see that if the organism persists in the direction indicated by the motive and goal, the "correct" response is hit upon and we have the first stages of *learning*—the acquiring of new responses or the modification of old responses. If we allow a period of time to elapse and then test the organism in the same situation, we are in effect making a test for *forgetting*. If there is continuous repetition of the correct response far beyond the point at which learning ceases, we have the situation which is commonly called *work*.

Figure 3, then, indicates the areas of study with which we will be concerned. Three working principles should be called to your attention since they indicate reasons for including and excluding certain material.

1. We will consider animal experiments an integral part of our study. There are three reasons for this: (*a*) many of our best-known theories of behavior are grounded at least in part on data obtained from animal experimentation; (*b*) a great many psychologists are engaged in research using animal subjects; (*c*) some of the finest experimentation, from the standpoint of careful control, is being done in the animal laboratories.

Lest you be misled you will find that you are consistently cautioned against making too hasty an application of the principles of animal behavior to the interpretation of human behavior, and vice-versa. This book will not, however, determine when you should, and when you should not, make such an application. We may only point out that in many instances there is a high similarity between principles of animal behavior and principles of human behavior.

2. No distinction will be made between so-called "applied" and "pure" experimental psychology. Sound methodology is always necessary to determine relationships between stimulus conditions and responses, and sound methodology makes no distinction between problems of a pure nature and problems of an applied nature. It is our purpose to give the initial training in sound experimental procedure; it is up to the individual to do what he will with that training. The illustrative experiments throughout the book will include certain "applied" experiments as well as experiments that are called "pure."

In a real sense the distinction between pure and applied science lies in the attitude of the working scientist. If he is doing work which has rather immediate social or practical application, and if this is the reason for his work, we would say he is working on applied problems. If a scientist cannot see such value to his work in the proximate future and continues to labor in spite of that, then he might be called a pure scientist. Whether an experiment is to be considered pure or applied must be determined by the attitude of the experimentalist at the time of its execution because human vision is so limited that work which we might call pure

today may, under certain conditions, become very much applied tomorrow. Of this the history of science shows many illustrations.

3. Some of the chapters in the book will be organized around consideration of the major stimulus variables which influence the process under consideration. There are certain variables inherent in the subject, such as age, sex, and height, which we shall not consider except in special instances. An experimenter cannot manipulate a subject's chronological age in the laboratory in the same manner that he can the subject's motivation. If he is interested in chronological age as it relates to a certain response variable, he will choose groups of subjects of different ages. In passing it may be noted that we shall not be concerned with individual differences to any appreciable extent. Both of these limitations are made because of lack of space, and not because they are unimportant problems. Our approach will be, in effect, *a dimensional analysis of manipulable variables.*

SCIENTIFIC ATTITUDE

We have already mentioned that nothing is more important for the beginning experimentalist than to develop actively a critical comprehension of experimental design. This ability, like the ability to handle the calculus, a foreign language, or a scalpel, can only be mastered if constantly practiced. It requires constant application to develop what has come to be known as the *scientific attitude.*

An exact analysis of the scientific attitude is a difficult undertaking. It seems at present that the best that can be done is to list some adjectives describing the behavior which seems to go with the attitude. In the first place, the scientific attitude is one of caution. It accepts alleged relationships only after some investigation of their derivation. In other words, method is all important. It is critical of generalized statements based on a small amount of data. It is skeptical of theory which is not anchored in empirical observations and is usually not concerned with theory which does not embody provisions for directing experimentation. Sometimes the scientific attitude has elements of cynicism in it, for it seems to evaluate the adequacy of experiments performed by "great names" with the same thoroughness as it evaluates those performed

by "unknowns." The scientific attitude is certainly imaginative. It is able to grasp relations and to propose experiments for testing theories and difficult problems. It is fortified with statistical techniques which force one to think in terms of the probability that an event will occur and not in terms of a dogmatic "will-happen" or "won't-happen" assertion. Finally, and above all, the scientific attitude is a pragmatic assumption that the fundamental laws of nature can best be determined by sound research fitted into a theoretical framework. Yet it should be understood that there can be no bias attached to experimental procedure as such, for certain basic principles of conduct are accepted as premises when one adopts the experimental method as a method for discovering the laws of nature. The reliability of experimental fact is then determined on the basis of how far, if at all, the experimenter has deviated from those rules for conducting experiments.

ODDS AND ENDS

1. As new concepts are introduced they will be defined. For ready reference a glossary of these concepts will be found at the back of the book.

2. The terms *subject* and *experimenter*, since they occur so frequently, will be abbreviated hereafter. Subject will be written, *S* plural, *Ss*. Experimenter will be written *E;* plural, *Es*.

SUMMARY

In this chapter we have introduced the basic concepts needed to discuss experimental work in psychology. Illustrations of controlled experimentation have been given to indicate some of the general factors which have to be considered in evaluating experimental designs.

We have discussed how experimental problems arise and have emphasized the importance of theory in suggesting experimental problems. Steps which experiments usually follow have been indicated. Considerable time was spent on the concept of dimension, because thinking in dimensional terms seems fundamental to the design of experiments for the exploration of the effect of variables.

The successive areas of study as they will appear in this book have been diagrammed by following through the sequence of possible events of an organism placed in a problem situation.

Finally, we discussed what seem to be some of the components of the scientific attitude.

CHAPTER II

Discriminal Processes: Methods of Study I

INTRODUCTION

We are able to respond in a selective fashion to the enormous complex of stimuli which impinge upon us at all times. We tend to take it for granted that because we are what we are certain stimuli will cause us to respond in one way, that other stimuli will bring about quite a different response, and that some apparent stimuli will elicit no noticeable response. The selective capacity of the organism, that is, the ability to respond in different ways to different stimuli, presupposes what we shall call the *discriminal processes*. The organism is constantly "exercising" these processes. Hence, such different reports as: "I'm tired," or, "I'm not tired"; "It is raining," or "It isn't raining"; "This is better than that"; "This coat is a darker brown than that one"; "That mountain peak must be about 30 miles away"; "These fence posts are out of line"; "I prefer Brahms to Debussy"; "Those two lights are of equal brightness."

For the most part, researches concerned with the discriminal processes have, for purposes of study, assumed these processes to be relatively unchanging. That is, little work has been done toward modifying these stimulus-response relationships so that the modification is relatively permanent. Cross-sectional analysis is the procedure wherein *E* attempts to determine the organization of the stimulus-response relationships as they already exist, with scant attention to the problem of how these relationships developed. By hundreds of such analyses, having their beginning even before Wundt established his laboratory in 1879, a multitude of stimulus conditions affecting the discriminal processes have been discovered. We shall not be able to treat these stimulus conditions

in a systematic manner. Rather, our major purpose will be to give a rather comprehensive consideration of the methods of experimentation which have been employed to study the discriminal processes, and we shall illustrate these methods with a wide sampling of studies to show the breadth of the problems to which the methods have been applied. Selective references at the end of Chapter IV will give sources in which a systematic treatment of the stimulus variables has been made.

E's job is that of relating variations in stimuli to variations in response. In studying the discriminal processes there are two distinct stimulus situations which have been used: (1) stimulus situations in which variation is described by physical dimensions; (2) situations in which variation in the stimuli cannot be meaningfully described by physical dimensions. Let us illustrate the difference. We present two stimuli to *S*, one a tone of 1000 cycles and the other of 1008 cycles. *S* is asked to judge which tone has the higher pitch. In this case we know the physical stimulus values —1000 and 1008 cycles—and we can relate the changes in cycles to changes in perceived pitch. On the other hand, we might present *S* with two modernistic paintings and ask him which of the two he prefers. These paintings rather defy description in terms of physical stimulus values. We have varied the stimulus situation, but we do not attempt to describe that variation in terms of physical dimensions. Such description may be meaningful in some cases, but in many it will not.

This and the following chapter will deal largely with situations of the first kind—situations in which stimuli are varied along a physical dimension. Methods used to study the stimulus-response relationships in these situations are commonly called ***psychophysical methods***. In Chapter IV we shall deal with situations of the second kind. As in most attempts at classification, we shall find it impossible to distinguish sharply among the methods as here classified. Titles or names of procedures should, therefore, be considered modal points around which one may organize his thinking, not mutually exclusive categories. Methods of experimentation will vary somewhat even though given the same name. We shall illustrate the "pure" method and then give samples of the procedure as it has been used in practice.

The two psychophysical methods to be considered in this

chapter are: (1) The method of average error, sometimes called the adjustment method; and (2) The method of limits, sometimes called the method of minimal changes.

THE METHOD OF AVERAGE ERROR

The Method

A housewife shows an interior decorator a sample of a very light pink color and asks that a room be painted that color. Since the color is not commercially produced, the decorator must mix paints to obtain it. Using standard white as a base, he very slowly adds red, stirring all the while, until the mixture appears to be a good match for the sample given him by the housewife. Unknowingly, the decorator has employed the method of average error as it would occur on a single trial in the laboratory.

FIG. 4. The Müller-Lyer Illusion. Line *A* appears to be shorter than Line *B* though both are $1\frac{1}{4}$ in. long. To determine the amount or extent of the illusion, apparatus is used in which the length of Line *B* is adjustable. *S* is asked to make several independent settings of Line *B* each time so that it appears to be equal in length to Line *A*.

The method of average error was designed to study the precision of observation or the precision of any matching procedure. If we state this in converse fashion, we would say that the method is used to study the errors which enter into observations. A good illustration of the method as it is applied in the laboratory is afforded by the well-known Müller-Lyer illusion experiment. Figure 4 shows this illusion. The differences in the direction of the "tails" tend to make Line B appear longer than Line A, even though physical measurements show they are exactly the same length. In using this illusion as a laboratory device, Line B is constructed so that it is variable, i.e., can be made longer or shorter. In determining the extent of the illusion, *S* is asked to make the length of Line B equivalent to the length of Line A. We would say, then, that *the method of average error consists in*

presenting S with some constant or standard stimulus and asking him to match it by manipulating a variable stimulus.

We may already suspect that a single match of the standard and variable would not be sufficient for a good index of the accuracy of observation. We will expect fluctuations or variations in *S*'s responses. Consequently, we will ask *S* to make several matches and then, after measuring the length of Line B for each setting, use a measure of central tendency of these lines—usually the mean—as the best estimate we have of the length of Line B which appears to be equal in length to Line A. *E* sometimes sets B so that it is obviously much longer than A, and sometimes he sets B so that it is obviously much shorter than A. The instructions to *S* in each case are that he should adjust the setting so that B is equal to A. The mean of these settings is the length of Line B which is subjectively equal to the length of Line A. We call the mean of the settings of B the length of subjective equality, although the generalized term for various kinds of discriminations is the *point of subjective equality*. If we put this concept into a formal definition we would say that *the point of subjective equality is that physical stimulus value of the variable which appears to be equal to the standard stimulus value.*

Due to the illusory nature of the figure, Line B will nearly always be set consistently shorter than Line A. This illusion is quite powerful and naïve *Ss* will be entirely unaware of the large errors they are making. The writer had constructed some of these illusions for use in the laboratory. The standard was 10 centimeters long. The construction was rather crude and the students were rather skeptical as to whether or not an illusion was being produced when they started the experiment. One *S*, for example, after having made several settings said in effect: "These things aren't any good—I don't get any illusion." An examination of the settings he had made, which he had not inspected on the data sheet, showed that his mean setting of B was 7 centimeters; on the average he had underestimated 3 centimeters or about 30 per cent of the total length of Line A.

This illusion provides an excellent laboratory exercise, since it affords a good initial illustration of stimulus variables that will cause the response to vary. It can be shown, for example, that the length of the "wings" or "tails" will influence the extent of the

illusion. Also the angle at which the "wings" intersect the horizontal lines will influence the amount of illusion. We would thus say that the angle of incidence is a stimulus dimension which influences the response and that the length of the "wings" is a stimulus dimension which influences the response, the response in each case being the extent of the illusion.

The geometrical illusions, of which the Müller-Lyer is only one illustration, provide simple examples of how our discriminal processes do *not* faithfully reflect the objective world. We do not always see things as they exist in physically measured reality. One of the best natural illusions is the common observation that the moon appears much larger when near the horizon than it does when directly above us. There is no general agreement as to the basic causes for many illusions. We do know that they will vary with changes in the stimulus conditions and that many of them can be reduced or even eliminated with continued practice. But, as we have already discussed, since we do not use stimulus variables as ultimate explanatory mechanisms, the fact that the illusions can be made to vary by changing conditions does not satisfactorily explain them. Boring (31) presents in brief form several of the theoretical explanations which have been offered for various geometrical illusions.

Constant errors. How much *S* errs in adjusting the variable stimulus of the Müller-Lyer illusion is best represented by the difference between the mean setting of Line B and the actual length of Line A. This difference is called a *constant error* and gives a quantitative measure of the extent of the illusion in this particular illustration. A constant error is produced by a uniform condition, or better, a uniform *set* of conditions that tends to "throw" the response in the same direction. In the Müller-Lyer illusion the constant error will produce underestimation of the length at which the variable should be set. Other situations may produce an error which will result in overestimation of the size to which the variable should be adjusted. It is possible that in certain situations some conditions may make for overestimation and others for underestimation so that the resulting point of subjective equality will be a function of the difference in the amount of influence of the two sets of conditions.

Every experiment, regardless of its simplicity, is subject to

possible constant errors, either because of the conditions of the experiment as such or because of the perceptual biases of *S*. In addition to the constant error produced by the illusion itself, the Müller-Lyer experiment has two other possible sources of error. One of these is called the *space error*, the other the *movement error*. The space error may be thought of as a bias of *S* for judging the variable differently when on the left, with the standard to the right, than when their positions are reversed. To determine if this error is operating in the case of a given *S*, or in the case of a group of *S*s, we would turn the apparatus over so that the variable Line B would be to the left of *S* half the time and to his right half the time. Separate records of the settings would be kept so that the amount of space error actually present could be determined. A space error is evident when the measurements with the variable to the left are significantly different from those obtained with the variable to the right.

The error of movement may be thought of as a bias which *S* may have for moving the variable outward as compared to moving it inward, or vice-versa. It is conceivable that different settings might be obtained if the variable is set shorter than the standard and then moved outward than when it is set longer and moved inward. To measure the movement error—or to determine if it is present—we plan the trials so that *S* is moving the variable inward half the time and outward half the time, and we keep a separate record of the results of these settings.

Even in so simple an experiment as the one we have been discussing there may still be other factors producing constant errors. As in most experiments, there may be practice effects, i.e. *S* becomes adjusted to the situation and a gradual shift in response may occur. Furthermore, suppose *S* gradually becomes fatigued or bored and this brings about a shift in response. Inasmuch as these factors may produce a systematic trend in the responses we must identify them as conditions producing constant errors. These two constant errors, practice effects and fatigue effects, *are potentially constant errors in almost all psychological experiments, hence, become factors which must be considered in the design of all experiments*. We must handle them in either of two ways: (1) distribute the effects so that they do not differentially influence various conditions of the experiment, or (2) design the experiment

so that the effects can be measured. These are problems best taken care of by counterbalancing.

Counterbalancing

Let us outline the conditions which are necessary to handle adequately the constant errors in the illusion experiment. To get stable results we normally would require *S* to make a great many settings, perhaps 100 or more. However, for illustrative purposes, let us restrict our consideration to a few trials; namely, 16, since that will divide properly for the measurement of the various constant errors involved. Any greater number of trials could be multiples of 16. Let *L* stand for having the variable stimulus to the left and *R* to the right. Let *O* stand for having the variable set shorter than the standard (*S* then moves it *O*ut), and let *I* stand for setting the variable longer and having *S* move it *I*n. The 16 trials would then be as follows:

L(4 trials)	*R*(4 trials)	*R*(4 trials)	*L*(4 trials)
O I I O	*O I I O*	*O I I O*	*O I I O*

Note that the basic sequence is *LRRL* and *OIIO* (this could be *RLLR* and *IOOI* as well), which when generalized is usually referred to as the *abba* sequence. The reason for this sequence is that if we have systematic errors which influence the response in a greater and greater amount as the experiment proceeds, their effects will not be distributed unequally on any of the conditions. We need to pursue this further.

Let us assume that systematic errors enter into the illusion experiment with ever increasing potency as the successive trials are run off. Suppose that instead of having the *L* and *R* settings *LRRL* as above, we had them *LLRR*, or simply, 8 trials *L* followed by 8 trials *R*. Under such conditions the effects of the systematic errors would be heavily concentrated on the *R* settings and would be of small consequence on the *L* settings which came early in the sequence. The assumption on which the *abba* order is based is that the progressive change in response which may take place as experimentation proceeds is a straight line function. This is illustrated in Fig. 5. The figure shows the hypothetical changes in the response which take place with successive trials as a consequence of the increased influence of factors making for

constant errors—such factors as fatigue, practice, and so forth. The abscissa (base line or X axis) indicates successive trials, 0 to n, the total of which is broken into four blocks and noted as *abba*. The ordinate (vertical or Y axis) is in hypothetical units indicating the amount of influence the constant errors have on the response we are measuring.

The average influence of the factors making for constant errors in the response during the first block of *a* trials is 1. To obtain this we assume that the first trial of the experiment—the first trial in the *a* block—is uninfluenced by systematic errors: Thus, a weight of *zero* is assigned. On the last trial in this *a* block 2 units of influence occur. Consequently, the average influence on all trials in the *a* block would be 1. Similarly, in the first block of *b* trials the amount of influence of the progressive error varies from 2 to 4 units with an average of 3. For each successive block of trials, *a b b a*, the change in response due to the constant error factors is 1, 3, 5, and 7 respectively. The amount of influence on the *a* conditions combined is 1 plus 7, or 8 units; the amount of influence on the *b* conditions is 3 plus 5, or 8 units. Thus, the amount of influence on both the *a* and *b* conditions is the same. This is our first illustration of a method of counterbalancing conditions. Its effect is that of distributing equally over all conditions the response changes produced by systematic or progressive errors.

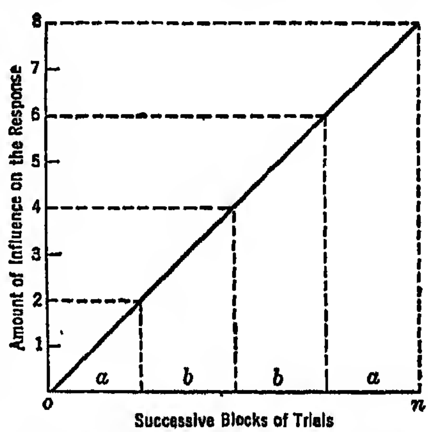


FIG. 5. How the influence of progressive errors on response can be equalized for all conditions by using the *abba* order of presentation. See text for complete explanation.

Now suppose we had not used the *abba* order, but had used instead, *aabb*. In such case the changes in the *a* conditions would amount to an average of 2 units while the changes in the *b* conditions would amount to 6 units. Thus, the responses would have

been differentially influenced by factors which were not the subject of study, i.e., not the stimulus variable. The systematic counterbalancing of conditions to remove the differential effects of constant errors when they are not objects of study is an extremely important procedure in experimental design and is one to which we will refer frequently.

It should be mentioned that the systematic trends in response due to practice effects, fatigue, or what not, will probably not follow a straight line function. Even when the function is not a straight line, however, the use of the *abba* order will not produce a serious bias in the results and is certainly much better than doing nothing about the progressive changes. This can be demonstrated to your own satisfaction by drawing in various plausible shapes of curves, assigning numbers to the Y axis as we have done in Fig. 5, and then averaging the *a's* and *b's* in the various instances. It will be seen that the differences which arise between the mean *a's* and mean *b's* will be much less than if presented in *aabb* order. We do not, of course, have to divide our total block of trials into four units. We could, and it is often wise to do so, repeat the *abba* order. For example, in our 16-trial illustration above, we might use *LRRLLRRL* with two trials given under each rather than four trials. Our *O's* and *I's* would be changed correspondingly. In general, the greater the number of *abba* sequences for a given series of trials, the less the possibility that any condition will be differentially influenced by progressive errors.

We cannot determine how much progressive change in response is due to fatigue and how much is due to practice effects, if any. All we can determine is whether or not there have been progressive changes in response as the successive settings were made. A comparison of the mean of the first eight settings with the mean of the second eight settings will show whether or not there was a change from the first half to the second half. We might go more to the extremes and compare only the first four settings with the last four settings since they have been obtained under identical conditions except for the temporal point in the series of settings. If the mean of the last group of settings is greater than the first set (if the illusion is less), we would say that there has been a lessening of the effect of the illusion with practice. If the mean setting of the last trials was less than the mean of the first trials (greater

illusion), we would say that there has been an increase in the effect of the illusion with an increase in the number of trials and then perhaps we might speculate that this is due to fatigue effects. Actually, the over-all trend which might be found could be a function of many factors, of which practice effects and fatigue are only two, and two very non-specific factors at that. We accept the fact that they are likely to occur and design all experiments accordingly.

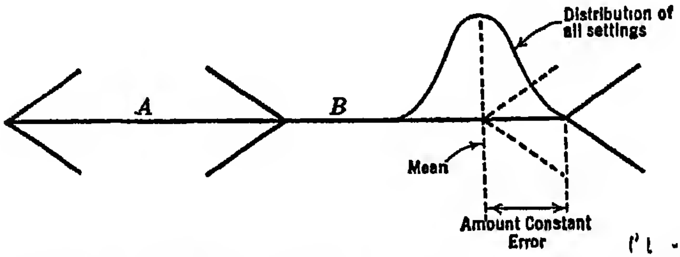


FIG. 6. Schematic representation of results of one S on the Müller-Lyer illusion. The intersection of the dotted "tails" on Line B indicates the length of Line B which on the average appeared to be equal to Line A. The distance from the dotted "tails" to the center "tails" indicates the length of subjective equality. The constant error is the difference between the length of subjective equality and the length which is objectively equal to Line A. The measure of variable error would be the standard deviation of the distribution of all settings.

Variable errors. We have discussed constant errors and have seen some of the possible sources of constant errors. It should be understood that the constant error is not "really" constant from trial to trial. In the illusion experiment, the illusory nature of the figure itself produces a constant error, but it does not produce an equal amount of error on each setting. The constant error is a statistical abstraction. What causes these fluctuations from trial to trial? The causes are not known specifically, and when a source of error is not known and does not appear to be in any systematic direction; we call it a *variable error.* These errors may represent variations in S, E, or in an instrument. Statistically we call them chance errors and usually measure them by the standard deviation or some other index of variability. In the same sense that we do not expect 12 tossed coins to constantly fall with six heads up and six tails up, we do not expect a given response to reproduce itself exactly from trial to trial.

In a final reference to the Müller-Lyer illusion experiment, we may diagram the hypothetical results from a single *S* for a great many settings. Figure 6 shows the constant error and the distribution of individual settings. The standard deviation of this distribution of settings may be used as the measure of variable error.

Illustrative Experiments Using the Method of Average Error

An experiment on direction perception. Let us suppose that we are standing some distance behind an anti-aircraft gun whose barrel is vertical. A plane towing a target approaches from the right so as to pass directly over the gun. As observers we probably would think that we could tell when the target was directly over the gun, i.e., that a bullet fired and taking no time to reach the target would hit the target. This is essentially the sort of problem attacked by Salomon (328). *E* was concerned with two problems in direction discrimination: (1) what are the errors (both constant and variable) involved, and (2) what are the stimulus dimensions which can be manipulated to increase and decrease these errors? In the laboratory the situation is imitated, of course, without the aid of flying airplanes and anti-aircraft guns.

E projected a line on a screen, tilted at about 20 degrees. Also projected was a small dot at some distance from the line. The projection equipment was so devised that *S*, by manipulating a knob, could move the dot at right angles to the upper end of the line. *S*'s problem was to set the dot so that the line, if extended, would pass through it. In our crude analogy above, *S* would have to move the target until it was directly over the line of direction of the gun barrel. The experimental problem is illustrated in Fig. 7. The series of unfilled circles shows how *S* could move the dot (filled circle) to the point judged to be the extension of the line.

Because our judgments of relations of objects in the environment are strongly influenced by other objects, it is desirable in the experimental situation to eliminate these objects unless we know their influence. Since we usually don't know their influence, *E* will eliminate them from the situation, thus reducing the number of variables which may be operating. This, in effect, makes the situation more "pure," and increases the probabilities that the

precise influence of a single stimulus variable can be determined. In technical language which has been applied to the perceptual situation, *E* is said to restrict the *frame of reference*. "The frame of reference may be defined as the background of stimulation which influences our behavior in a particular situation" (55, p. 17). If *E* restricts the frame of reference he is restricting the number of stimuli which may influence the response. Almost all laboratory situations modify the frame of reference by making it more homogeneous than situations in which the phenomenon occurs in "real life."

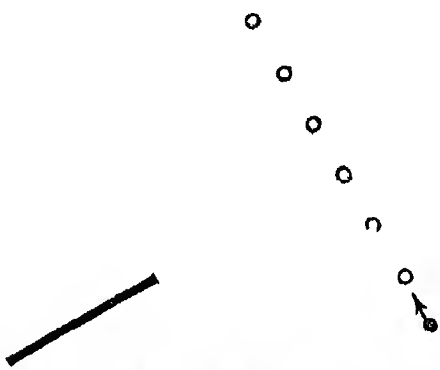


FIG. 7. Illustration of the situation used by Salomon (328) in an experiment on direction perception. A tilted line and a dot (filled circle) were projected on a screen. The dot could be moved in either direction in the plane indicated by the unfilled circles. *S* attempted to set the dot so that an extension of the line would pass through it.

In the experiment under consideration, *E* markedly restricted the frame of reference. The line and the dot were projected on a perfectly homogeneous white wall. There were no smears, no nails, nothing on the wall that would provide *S* with orienting cues. *S* sat in a small dark room and observed the line and dot through a circular opening 25 inches in diameter. Seated 3 feet from this opening, with his head held rather firmly by a padded steel frame, *S* could see nothing but the white background and the projected line and dot. The screen (white wall) was actually 15 feet from *S*, but because of the lack of cues it was extremely difficult for *S* to tell how far he was from the screen. To have *S* looking through the circular opening at an homogeneous background is an application of the *reduction screen* technique for

limiting the frame of reference. The reduction screen is usually used on a smaller scale, in that *S*'s eye is directly in front of a very small hole through which he can see only a limited area.

Before running the main experiment, *E* tried out several procedures to determine the most adequate method. This is often necessary when one attempts experimentation in a relatively new area or with new apparatus. These initial or preliminary studies are often called "pilot" or "test-tube" studies. We may wonder, for example, how it came about that Salomon used a circular opening. In pilot studies Salomon used both a circular and a rectangular opening. Finding that the judgments in the two instances did not differ significantly, i.e., that the shape of the opening was not a stimulus variable in this situation, she chose the circular opening because *S*s seemed to prefer it. In the main experiment *E* used several lengths of lines projected on the screen since it was found that this was a major stimulus variable influencing the judgments. How did the length of lines used happen to be chosen? In preliminary work it was found that a line $^1/_4$ inch long was the shortest length which yielded consistent observations. It was thus determined that lines shorter than this would produce no useful results. In the main experiment, to be on the conservative side, the shortest line used was $^1/_2$ inch.

In the illusion experiment we discussed the possibility of movement error. We have a comparable situation here. The dot could be set clearly below the direction point indicated by the line or clearly above, with *S* moving it up and down respectively. Preliminary experiments showed that there was no constant error involved in this task; hence, the design of the experiment did not have to account for that. These exemplify the preliminary findings used in determining the conditions for the main experiment.

In the major experiment Salomon varied two stimulus dimensions: (1) the length of the line; and (2) the distance of the dot from the line. The line lengths used were four in number, .5, .75, 11, and 22 inches. Three distances of the dot from the end of the line were used, these being .25, 2.5, and 5 feet. From this design, 12 different empirical points, indicating 12 different combinations of line length and dot distance, could be determined for each *S*.

For each 12 combinations of settings, 40 trials were given. Each of the 10 *S*s thus made a total of 40 by 12, or 480 settings. Since

progressive changes in the response may take place as successive trials are run—changes due to fatigue, practice, and so forth, as already discussed—all *S*s did not follow the same order of conditions. As in the illusion experiment, *E* distributed the progressive changes as equally as possible over all combinations of conditions. To accomplish this a random order of conditions was used. One *S* made his first 40 settings with a line of .5 inch and the dot .25 feet away; another *S* made his first 40 settings with a 22-inch line and the dot 5 feet away, and so forth, so that for all data for all *S*s no combination of conditions was favored by appearing early or late in the experiment. *The use of a random order of conditions for Ss is thus another way in which practice and fatigue effects may be equalized for all conditions.*

The following instructions were given to *S*:

> This is an experiment in the perception of direction. A line, like this, will be projected on the screen, and a dot a certain distance away is to be placed in the direction the line is pointing. (You can imagine it's a gun shooting a bullet—put the dot where the bullet would hit it.)
>
> You will be sitting in this booth, with your chin on the head rest, and your left arm reaching out to this handle which moves the dot up and down. It moves about at right angles to the line. You can use the handle till you get the dot close to the right place, and then use the edge of the dial for the fine adjustment.
>
> The purpose of the head rest is to keep your head horizontal, and to keep you from looking too far out and seeing the corners of the room. You can move your *eyes* as you want to, as long as you keep your head still. Set the dot by any method you want to. There is no time limit, but it is best to set it where you think it should be, where it looks right, quite quickly, and leave it there. Sometimes if you deliberate and think about it too much, the setting is not as good as the first impression. (328, p. 73)

After three or four preliminary practice trials the experiment proper got under way. Note from the instructions that it is necessary for *S* to manipulate a dial and vary the position of the dot. However, this dial was so constructed that *S* could get no cues as to where he had set the dial on the preceding trial. Thus, there was no possibility of *S*'s simply setting the dial itself at the same place trial after trial without regard to the line and dot. From another calibrated dial hidden from *S*, *E* recorded the setting made by *S* on each trial. This was transferred at the same time into inches of error (or fractions thereof). *E* was able to determine

what constituted a perfect setting through the aid of a separate device; hence, S's settings could be transferred quickly into a deviation-from-the-perfect score.

TABLE 1

VARIABLE ERROR IN THE PERCEPTION OF DIRECTION

Distance (ft.)	*Length of Line (in.)*			
	.5	.75	11	22
0.25	.14 (.02)	.17 (.04)	.09 (.02)	.10 (.03)
2.50	.92 (.14)	.74 (.07)	.58 (.08)	.52 (.10)
5.00	1.79 (.11)	1.56 (.14)	1.18 (.12)	.86 (.10)

The data express the relationship between the standard deviation (variable error) and the length of line and distance of dot from the line. The figures indicate the mean standard deviation of 10 *S*s. The figures in parentheses are standard errors of the standard deviations. (1.00 equals .75 inch displacement on the screen). Data from Salomon (328).

The results of this discrimination of direction experiment, as regards the variable error, are presented in Table 1 and Fig. 8. The table provides the detailed quantitative results and the graph will give a quick understanding of the basic relationships. From the table and the figure, two general relationships can be stated: (1) The greater the distance of the dot from the end of the line the less the precision of the response (the greater the variable error); and (2) the shorter the line the less the precision of response.

So much for the variable error. Did *S*s show any constant error? Was there evidence of any bias for consistently setting the dot higher or lower than the true projection of the line? At short distances there was little evidence for such a constant error, but when the dot was 5 feet from the end of the line (regardless of the length of the line), the dot was placed consistently below the true extension of the line. No cause for this bias is immediately discernible.

During the main experiment which we have reviewed, *S*s were never told how accurate their settings were. As we shall see later, knowledge of performance is an extremely important factor in most human learning. Supposing that in the present experiment *S* was informed after each trial the direction and extent of his

error. We would probably predict better performance as a consequence of this procedure. In a minor experiment, Salomon did this very thing and found that the precision of response was increased considerably as a consequence of this knowledge of performance. This increased precision was most marked with long lines and short distances between the dot and line. Furthermore, the constant error which was present for the greater distances was greatly reduced.

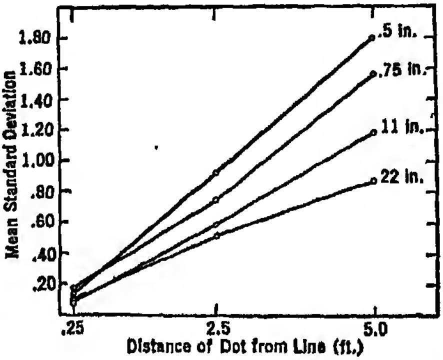


FIG. 8. Variable error in direction perception. Two relationships are shown: (1) distance of dot from line (abscissa) and variable error (ordinate), and (2) length of line (as noted) and variable error. Data from Salomon (328).

Salomon found (1) that as the dot was placed farther and farther from a line of constant length, variability of response increased; and (2) that as the line decreased in length, variability increased. Thus, two stimulus dimensions have been explored and the precision of the discriminal responses involved determined. We do not conclude that these are the only two effective stimulus variables in the situation. We can surely suggest others which would probably influence the response—such factors as the distance of *S* from the screen, size of the dot, width of the line, the amount of lighting on the screen. We can see that even in this relatively simple discrimination situation it takes a long series of researches to determine (1) which variables are effective in producing changes in response, and (2) precisely what those changes are as the stimulus dimensions are varied from low to high. The

collecting of complete data on a given response requires a systematic exploration of many stimulus dimensions.

TABLE 2

RELATIONSHIP BETWEEN LENGTH OF LINE AND CONSTANT AND VARIABLE ERRORS IN THE LINE-DRAWING EXPERIMENT.

Deviation from Standard	*Length of Standard Line*		
	16/16 in.	32/16 in.	48/16 in.
— 3	0	0	2
— 2	6	1	5
— 1	14	5	6
0	56	9	11
1	110	18	16
2	126	46	38
3	87	76	65
4	33	87	74
5	8	76	60
6		66	64
7		36	45
8		15	31
9		5	13
10			8
11			2
Mean length of reproductions	17.747	36.254	52.606
Standard	16.000	32.000	48.000
Constant Error	+ 1.747	+4.254	+4.606
Standard Deviation	1.37	1.99	2.49

Line-drawing experiment. There are many experiments conducted by the method of average error or slight modifications thereof which demonstrate how the precision of the different discriminal process varies with certain stimulus conditions. The so-called line-drawing experiment is a simple one and clearly indicates changes in the variable error as a function of a stimulus condition. *E* took three 5 inch by 8 inch cards and on one drew a straight black line 1 inch long; on another a 2-inch line, and on the third, a 3-inch line. These three cards were the standard stimuli for the experiment. *S* was given one card at a time and instructed to place it upright—with the line horizontal—at the top

of a sheet of unruled paper. *S* was then instructed to attempt to reproduce the length of this standard line with a single sweep of the pencil. No special sighting was allowed in making the reproductions and *S* covered all preceding lines so that each new line was drawn on an homogeneous field. No knowledge of results was given.

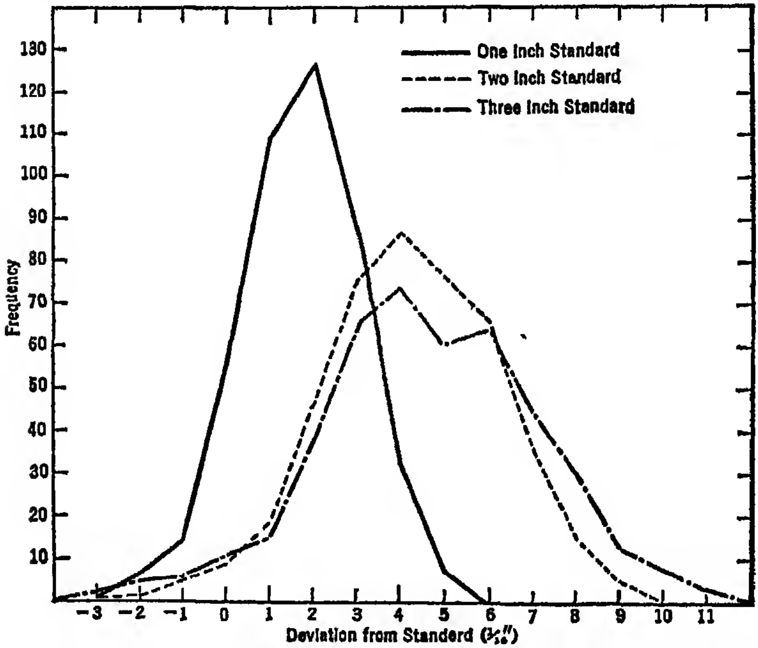


FIG. 9. Results of line drawing experiment. Three standards were used as indicated and *S* made 440 reproductions of each. Deviation from standard is plotted against the frequency with which that deviation occurred.

S made 440 reproductions of each standard for a total of 1320 lines. These reproductions were distributed almost equally over 12 experimental sessions so that, on the average, 110 lines were drawn during each session. An equal number of lines of different standard lengths was drawn at each session and a random order of presenting the standards was followed within each session.

In calculating the results it was first determined how much *S* deviated from the standard on each attempted reproduction.

Measurements were made to the nearest 1/16 inch. For each standard there would be a distribution of 440 deviations from the standard. If the mean of these deviations is added to the standard it gives the length of subjective equality, while the mean itself indicates the direction and extent of constant error.

Table 2 and Fig. 9 show the data obtained. Two facts are evident from the data on this one *S*: (1) As the length of the standard line becomes greater, the variable error becomes greater. This is shown by the increased range in Fig. 9 and the increased standard deviation in Table 2. (2) There is a constant error toward overestimation of the lines.

Movement perception. A brief preliminary consideration of certain visual phenomena will aid in understanding the illustrative experiment to be outlined in this section. Theoretically, any reliably measured constant error might be considered indicative of the illusory nature of the percept providing the constant error is based on relative judgments and not on absolute judgments. That is, if we asked *S* to estimate the absolute length of a series of lines we would very likely find a constant error, but we would not call this an illusion as we have in the case of the Müller-Lyer figure where the variable is compared relative to a standard.

One interesting illusion is shown by fixating steadily a weak light source in an otherwise darkened room. After a short fixation period the light will appear to move. It may appear to move in any direction and most people are subject to the illusion. This phenomenon has been called the *autokinetic illusion*, and is an illustration of *apparent movement*—movement which objectively does not take place. There are many other illustrations of apparent movement. Neon signs which give the impression of movement actually show "stills" in quick succession. Without apparent movement Hollywood would not exist, since movies depend upon rapid presentation of successive still pictures to give the illusion of movement. The autokinetic illusion may be the cause for some airplane accidents. Flying formation on a dark night with only the weak tail light on the plane in front as a guide to spatial orientation may be a near perfect set-up for the autokinetic effect if the pilot does not continually break his fixation. Other phenomena of illusory character may be present during flight. Angular acceleration (change in rate of rotation) and *g* forces (as

exemplified by the outward pull on a merry-go-round or as felt in centrifugal swing) may each cause apparent movement of lights which are actually stationary (121). The true importance of these illusions for aviation safety must await further study.

We may have distortion in the perception of real movement. This will be illustrated by one of a series of experiments by Ansbacher (5) in which the method of average error was used.

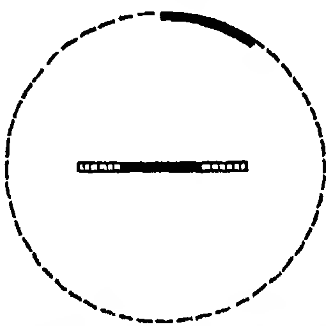


FIG. 10. Essentials of the stimulus situations used by Ansbacher (5) in his study of distortion in perception of real movement. The arc-shaped darkened slit represents a slit of light which could be rotated at various speeds around the circumference of the circle. The length of this arc was held constant. The slit of light in the center of the circle was variable as indicated by the dotted extensions. *S* was to match the length of the slit in the center with that of the arc slit.

The device used by Ansbacher is illustrated in Fig. 10. The two essential stimuli were slits of light, one, the standard stimulus, being an arc of light of 36 degrees—1/10 of the circumference of the circle of which it was a part. This arc could be made to rotate around the outlines of the circle at various speeds. The variable stimulus, a straight horizontal slit of light with a dot in the center, was located in the middle of the circle. The following instructions were given *S*:

You will see two illuminated figures: one stationary line here in the center, the length of which can be varied, and some other design rotating around the periphery of the disc in front of you. Look at the mark in the center of the stationary line. The line will at one time be very short at the beginning and expand slowly, the next time it will be long at the beginning and contract slowly. When the stationary line which is slowly being lengthened or shortened seems to you to be equal in length to the line or

figure you see moving around it, say "Now." Make judgments only when your eyes are fixated on the center point of the line. Make quick judgments; do not deliberate in any way. We are not interested in how correctly you judge the actual length of the figure; we only want to know how long the figure seems to you. Everything will be explained to you at the end of the experiment. (5, p. 4)

The device was designed so that the wheel would be rotated at different speeds. In the experiment, five different speeds were used; 0, 0.5, 0.7, 1.0 and 1.3 revolutions per second. Each *S* made 10 judgments at each speed, and all 10 judgments for a given speed were made before moving on to judgments at another speed. Of the 10 judgments in each case, half were made with the variable set much longer than the standard followed by gradual shortening, and the other half were made with the variable set much shorter than the standard and then gradually lengthened. Six of 12 *Ss* started with the zero speed (judgments made with no movement of arc slit) and proceeded up the five speed steps. The other 6 *Ss* started with the top speed and worked down. As Table 3 demonstrates, the two groups of *Ss* produced different results, which may be a function either of differences in the perceptual ability of the groups or of the order in which the judgments were made.

TABLE 3

ERRORS IN THE PERCEPTION OF REAL MOVEMENT

	Speed of Standard (r.p.s.)				
	0	0.5	0.7	1.0	1.3
6 *Ss* starting with zero r.p.s.	10.2	8.1	7.6	5.9	3.9
6 *Ss* starting with 1.3 r.p.s.	12.3	11.9	9.1	4.5	1.5
All *Ss* combined	11.3	10.0	8.4	5.2	2.7

The standard or arc-shaped stimulus was 13 centimeters long. The data show that this length was consistently underestimated in setting a variable stimulus so that the length subjectively matched the standard. Data from Ansbacher (5).

Two facts stand out from inspecting the mean judgments in Table 3. It can be seen that *Ss* underestimated the length of the arc even when the arc was not moving. Furthermore, the data

are clear in showing that the underestimation increases directly as the speed of movement around the circle increases. This is an illustration of distortion in the perception of real movement and is to be contrasted with apparent movement.

Summary of Section on the Method of Average Error

1. In its basic form the method consists in presenting some standard stimulus to *S* and requesting that he match or reproduce this stimulus by manipulating another or variable stimulus. The method has been largely used to determine the precision of the discriminal processes. Our examples of the method have been from the field of vision, but the method is adaptable to discriminal processes for most sense modalities.

2. Such researches have resulted in the identification of certain stituations which produce consistent overestimation or consistent underestimation of the magnitude of the standard. These errors in observation are called constant errors.

3. In addition to constant errors which may be a function of the stimulus as such, other systematic response changes may be brought about by practice, fatigue, and perhaps other factors. Unless these factors are the object of study, *E* must design his experiment so that they will not differentially influence the conditions.

4. Successive discriminations in the same situation show that the response fluctuates from trial to trial. For any given discriminal process, then, we expect a distribution of measurements indicating these differences from moment to moment. These distributions are evidence for what are called variable errors, the causes for which are not definitely specified. We may use the *standard deviation of the distribution of measurements* as an index of the size of the variable error. The variable error will change as a function of a great many conditions for any given situation, the study of this variation being one of the prime purposes of the method of average error.

5. The errors of observation are not only topics of study in and of themselves but must necessarily concern all scientists who continually make observations. The less the variable error the higher the precision of prediction from the given set of conditions.

THE METHOD OF LIMITS

Thresholds

The method of limits has been used primarily to determine thresholds (limens). There are two sensory thresholds which have been defined: (1) the absolute threshold; and (2) the difference threshold. Let us illustrate these concepts before we give operational definitions, i.e., before we define them in terms of the methods used to measure them.

Though the absolute threshold is sometimes called a stimulus threshold it might even better be called a *threshold of response.* Any energy to which the organism is sensitive must exist in a certain amount before it will arouse a response—before *S* will report the presence of a stimulus. The lowest stimulus value which will elicit a response is said to lie at the lower absolute threshold. For most people 16 to 20 cycles per second represents the lower threshold for sound. Frequencies below this are not heard. Since very high-frequency tones—approximately 20,000 cycles or above—can seldom be heard by most people, we can by experiment determine the upper absolute or terminal threshold. The range of stimuli from the lower absolute threshold to the upper absolute threshold for all characteristics or attributes of energies to which the organism is sensitive would define the stimuli to which one might expect a response.

Individuals taking the physical examination for entering the armed forces were often given the "whisper test" for determining auditory sensitivity. The examiner had presumably built up a series of whispers of different intensities. At a specified distance from *S* he knew approximately what the normal lower absolute threshold should be. Then to determine if there were wide deviations from the normal, he crudely explored *S*'s ability to hear the whispers. For careful testing an audiometer (a device for presenting tones of known intensity and frequency) is used.

Having heard a tone with a certain known frequency of vibration, how large an increase or decrease in frequency will be necessary before *S* will signify the pitch has changed? The amount of physical change necessary to bring this about is called the *difference threshold*—a just noticeable difference. Let us take another illustration. Given a standard weight of 100 grams, how many

grams will have to be added to that standard before S will indicate that the new stimulus is heavier than the standard? This again would be the amount of physical change necessary to produce a difference in the "felt" weight. To use an everyday illustration, how much weight must a 100-pound woman gain before it becomes noticeable to her friends?

Let us now be a little more precise with our definitions of the two thresholds:

Absolute threshold: That minimal physical stimulus value (or maximal for upper thresholds) which will produce a response 50 per cent of the time.

Difference threshold: That physical stimulus difference that is noticeable 50 per cent of the time.

As the above definitions indicate, we will have to apply statistics to get an adequate value for a threshold in any given situation. Thresholds will vary with individuals and they will also vary from *moment to moment* for a single individual. The measured absolute threshold or difference threshold on one trial will probably be somewhat different from the value on the next trial. The best measure we can provide of a threshold is a statistical abstraction—the mean or median of many threshold measurements. These may be taken on the same S if we wish to define his threshold, or we may take several measurements on many Ss if we wish to define a generalized threshold.

One word of caution should be given concerning statistical abstractions. Considering a mean value as a measure of a threshold, students will occasionally fall into the error of wondering: "But, what is the 'real' threshold?" This, of course, is an unanswerable question. We can know phenomena only by the ways in which we measure or observe them, and for all scientific purposes the statistical threshold is therefore the "real" threshold. Johnson (183) tells of a study in which a group of speech experts judged the number of times a so-called stutterer stuttered. There was a wide divergence among these experts in their tabulations. Johnson indicates that after reporting this study to a class he will occasionally get the query: "But, how many times *did* the person stutter?"

Most students do not have difficulty conceptualizing a threshold, whether it be a stimulus threshold or a difference threshold.

Numerous personal experiences aid one to ground the concepts. At one time or another most people have viewed a star that was barely visible and have noticed that it seems to appear, then disappear, and then reappear. A ticking watch held at a certain distance from the ear will alternately be heard and then not heard; the ticking appears to wax and wane. At one moment you may be able to distinguish a difference in two shades of gray, whereas at the next moment the same two stimuli will yield no perceptible difference. The difference threshold varies in much the same fashion as the absolute threshold. Around the absolute threshold and at the difference threshold as we know them statistically there is truly an area of uncertainty which appears to expand and contract from time to time.

The Method

Lower stimulus thresholds. In a completely darkened room we have installed apparatus to measure your lower absolute threshold for light intensity. The apparatus allows us to increase or decrease the intensity of the light source so that it can be made clearly visible or not visible. (Actually, this apparatus would be extremely complex. In addition, *S* would have to be thoroughly dark adapted by remaining in the dark room at least 30 minutes before the experiment started. A fixation point consisting of a dull red light would be provided). In using this apparatus to determine your lower absolute threshold for light intensity, we may, (1) start the procedure by making the light plainly visible to you and then gradually reduce the intensity until you inform us that you can no longer see the light, or (2) we may start with the stimulus value of the light definitely below your threshold of visibility and then gradually increase the intensity until you report seeing it. Actually, we would give you both conditions, repeating these descending and ascending series a good many times in alternate order. For each ascending and descending series we would get a threshold measurement for you—momentary as it may be. We would get some composite of the measurements to give the best estimate of your lower absolute visual threshold under these conditions.

Table 4 is a representation of ten series of measurements, five ascending and five descending, as they appear on the raw data

sheet prepared by *E* (cf. 422). Let us say that the intensity scale is in arbitrary numerical units rather than in terms of foot candles or millilamberts as would be the actual physical scale. In the table a *plus* indicates the light source as being reported visible by *S*, a *minus* indicates that it was reported not visible.

TABLE 4

ILLUSTRATION OF A SERIES OF MEASUREMENTS TO DETERMINE THE LOWER ABSOLUTE THRESHOLD TO LIGHT

Stimulus Value	Asc.	Des.	Asc.	Des.	Asc.	Des.	Asc.	Des.	Asc.	Des.	Stimulus Value
20								+			20
19								+			19
18		+						+		+	18
17		+				+		+		+	17
16		+				+		+		+	16
15		+		+		+		+		+	15
14		+		+		+		+		+	14
13		+		+		+		+		+	13
12		+		+	+	−		−		+	12
11	+	+		−	−				+	−	11
10	−	+	+		−				−		10
9	−	−	−		−		+		−		9
8	−		−		−		−		−		8
7	−		−		−		−		−		7
6	−		−		−		−		−		6
5	−				−		−				5
4	−						−				4
3	−										3
Individual Thresholds	10.5	9.5	9.5	11.5	11.5	12.5	8.5	12.5	10.5	11.5	

Mean Threshold Value: 10.8 Standard Deviation: 1.27

The transition point in each series in Table 4 is indicated by the step interval bounded by the points where *S* reports that he sees the light and then doesn't see it, or vice-versa for the ascending series. Where is the threshold in each series? Certainly not at the point where *S* still sees the stimulus, and just as certainly not at the point where he no longer sees it. For each series the best

estimate of the threshold is some value between the points "see" and "don't see." For example, on the first series it must be somewhere between 10 and 11 and the best guess we can make of it would be halfway between—10.5. On the next series it would be 9.5, and so forth. The mean of the individual thresholds would define our accepted absolute threshold value—that stimulus value which will elicit a response 50 per cent of the time. Note that once a change has been reported by *S* no further observations are made in that series, i.e., in the ascending series we never go beyond the first report of "see" (plus), and on the descending series never beyond the first report of "don't see" (minus).

As in the method of average error, the standard deviation is usually used as a measure of variable error. Do we have conditions which make for constant errors in the measurement of thresholds? There are such conditions, two of which will be considered here. These are, of course, in addition to the ever-present possibility of progressive changes in response brought on by fatigue and practice effects as previously discussed.

One of the constant errors which enters into the method of limits is called the *error of habituation.* Let us suppose that in a descending series we have our light well above threshold and then gradually reduce the intensity. The error of habituation supposes that *S* will fall into a "habit" or "set" toward giving the response "see," and thus continue reporting this past the point where the stimulus became invisible. The error of habituation would thus tend to make the descending series threshold lower than the ascending series threshold.

The other error, the *error of expectation,* is an error which is thought to be introduced because *S,* expecting a change, will be suggestible enough so that he reports a change before one is actually apparent. Such an error would tend to make the ascending thresholds lower and the descending thresholds higher. Upon reflection it will be noted that the influence of these two constant errors is directly opposed. If the influence of the two errors is the same—if the habituation tendency is about the same strength as the expectation tendency in a given *S*—no bias would actually be measured in the data since in combining the data a cancellation effect would prevail. If the effect of one is somewhat stronger than the other it should show up in the data. However, this effect

can be determined only by comparing ascending and descending series—not by analyzing a single trial or a group of ascending or descending trials. It is left as an exercise for the student to determine how the influence of each of these errors could be measured from data such as shown in Table 4. Both errors can be minimized by careful instruction to *S*, and by varying the level at which each successive series is started. Thus, *S* does not get "set" for any particular number of stimuli before a change, nor is he likely to become habituated to the sequences. It will be noted that in Table 4 the starting point for each successive series varies.

Difference thresholds. Let us consider next the application of the method of limits to the measurement of difference thresholds. We will continue with the visual problem, and add another light source to our apparatus. This new light source is used as a standard stimulus, the intensity of which will not vary. We set this standard at a known physical intensity value and then proceed to find out how much the variable stimulus must differ from the standard before *S* reports a just noticeable difference. Again we would use the alternate ascending and descending series. We start the variable light at an intensity value that is clearly brighter than the standard (descending), or at a point that is clearly less bright than the standard (ascending), and then gradually decrease (or increase) the intensity of it. As the intensity is gradually decreased, a point is reached at which *S* reports the two lights are of equal intensity. We may not stop at this point, but instead we continue decreasing the intensity of the variable until *S* reports that the variable is now perceptibly less intense than the standard. We have taken *S* through successive experiences of "brighter," "equal," and "less bright." The procedure would then be repeated with the ascending series, with the variable light being set initially so that it is clearly less intense than the standard, and then gradually increasing its intensity.

Data from such an experiment would appear as in Table 5. The standard stimulus was set at a value of 50. First we have an ascending series in which the variable is set lower than the standard and then its intensity gradually increased. *S* reports when the variable becomes equal to the standard in intensity and then reports further when the variable becomes brighter than the standard. In the descending series the procedure is reversed.

TABLE 5

AN ILLUSTRATION OF THE MEASUREMENT OF DIFFERENCE THRESHOLDS BY THE METHOD OF LIMITS

Stimulus Value	Asc.	Des.	Asc.	Des.	Asc.	Des.	Asc.	Des.	Asc.	Des.	Stimulus Value
58				+						+	58
57		+		+				+		+	57
56		+		+		+		+		+	56
55		+		+		+		+		+	55
54		+	+	+		+		+	+	+	54
53	+	+	=	+	+	=		+	=	+	53
52	=	+	=	+	=	=		=	=	+	52
51	=	=	=	+	=	=	+	=	=	+	51
50	=	=	=	=	=	=	=	=	=	=	50
49	=	=	=	=	=	=	−	=	=	=	49
48	−	=	−	=	−	=	−	−	=	−	48
47	−	−	−	=	−	−	−		−		47
46	−		−	−	−		−		−		46
45	−		−		−		−		−		45
44	−		−		−		−		−		44
43	−				−				−		43
42					−						42
Upper Threshold	52.5	51.5	53.5	50.5	52.5	53.5	50.5	52.5	53.5	50.5	
Lower Threshold	48.5	47.5	48.5	46.5	48.5	47.5	49.5	48.5	47.5	48.5	

Mean Upper Threshold: 52.1
Mean Lower Threshold: 48.1

Upper difference Threshold: 52.1 − 50 = 2.1
Lower difference Threshold: 50 − 48.1 = 1.9

For each series it will be evident that there are two transition points and that we may speak of thresholds of change. First, in the ascending series, there is a change from "less than "to "equal." This may be called the lower threshold of change. Secondly, there is a change from "equal to" to "greater than" the standard. We thus have two difference thresholds based on the two thresholds of change. In computing the actual values of these thresholds we take the difference between the standard stimulus value and the

stimulus value for the threshold of change. The threshold of change would be taken as that value which lies halfway between the two values marking the change. For example, in the first series the last *minus* is at 48 and the first *equal* judgment is at 49. Hence, the lower threshold of change is at 48.5. The last *equal* is at 52 and the first *plus* is at 53, so the upper threshold of change is 52.5. Taking the difference between the standard stimulus and the upper threshold of change would give us the upper difference threshold or a just noticeable difference value. Taking the difference between the standard stimulus and the lower threshold of change would give us a lower difference threshold. Thus we have for each series an upper difference threshold and a lower difference threshold.

If there is a constant error involved in the determination of these difference thresholds, we would not be justified in using the physical value of 50 to determine the difference thresholds. By so doing we would get a considerable discrepancy between the size of the upper and lower difference thresholds. If the constant error is large, i.e., if the point of subjective equality differs significantly from 50 in the sample problem, we would use this point of subjective equality to determine the difference thresholds. The point of subjective equality may be determined by taking the mean of all equal judgments. For example, the best measure we have of the point of subjective equality for the first series is 50.5—midway in the band of equal judgments. For the second series it would be 49.5, and so forth. The mean of these points would give us the best measure of the point of subjective equality, in this illustration being 50.1. If this point differs significantly from 50 we would not be justified in using 50 as the standard for determining difference thresholds. Suppose, for example, that the point of subjective equality was 52 instead of 50.1. If this were true and if we used 50 as the value for determing both the upper and lower difference thresholds, we would find that the lower difference threshold would be considerably smaller than the upper difference threshold. In calculating the difference thresholds, then, it is wise to determine first the point of subjective equality, and if that value differs significantly from the standard physical stimulus value, the value of the point of subjective equality should be used for determining differnce thresholds.

Weber's Law

We have considered in some detail the mechanics and rationale for finding a difference threshold. It can be seen that our end product is the amount of change in the stimulus necessary for *S* to note the change 50 per cent of the time. Now, this is not a constant amount of change for different values of any one physical stimulus dimension. There is a fairly general principle which governs the relationship between the amount of standard stimulation and the amount of change from that standard necessary to bring about a just noticeable difference. This principle of stimulus-response relationships in its simple form is known as *Weber's Law.* In the preceding problem on the difference thresholds, the data were set up so that the difference threshold below the standard would be slightly smaller than the difference threshold above the standard. If the measurements are precise such a slight difference would be found. Now let us exaggerate the situation. Suppose we had a very weak standard stimulus—just slightly above the lower absolute threshold—and determined a difference threshold at this point. Then suppose we used a very strong standard stimulus on the same dimension and determined another difference threshold. At these two extreme points will there be a difference in the amount of change in the physical stimulus necessary to produce a just noticeable difference? If you will check against your everyday observations you will answer in the affirmative. It is a commonplace observation that in turning on a 3-way lamp in which the successive intensities are 100, 200, and 300 watts, the change in the brightness of the room appears to be much greater when changing from 100 to 200 watts than it does in changing from 200 to 300 watts. Pouring a pint of water into a pail which has only a small amount of water already in it will produce a very noticeable difference in the amount of water in the pail. Pouring a pint of water into your city's storage basin will bring no perceptible change in the amount of total water. Add five students to a class of 1000 and compare that with the "felt" increment of adding five students to a class of ten. Clearly, these observations suggest that the amount of change in the physical stimulus necessary to produce a given amount of perceived change is some function of the amount or size of the standard stimulus.

Furthermore, these observations suggest that the stronger or larger the standard stimulus the greater the increment necessary to produce a perceived change. This relationship holds in general throughout the wide range of stimuli to which the organism is sensitive.

For quantification of the general principle we need to know the consistent or systematic relation between the value of the standard and the amount of change in that standard necessary to produce a just noticeable difference. Weber's law is one statement of this mathematical relationship. The law states that *the just noticeable increment to any stimulus bears a constant ratio to that stimulus.* The just noticeable increment would be a difference threshold. The ratio, of course, varies with different observers and different sense modalities, and with different dimensions along which a stimulus may be varied within a modality. Suppose we had 100 candles burning in an otherwise darkened room, and that the addition of one candle would produce a just noticeable increment in brightness. If we had 1000 candles burining we would have to add 10 candles to produce a perceived difference and with 10,000 candles an addition of 100 would be necessary. If we put Weber's law into a verbal equation we would have:

$$\frac{\text{Difference Threshold}}{\text{Standard Stimulus}} = \text{Constant}$$

If we vary the amount of the standard stimulus along a given dimension and measure the difference threshold at each point of change, the ratio of the size of the standard and the size of the difference threshold should remain about constant. In a wide variety of tests it has been shown that the equation holds roughly. The precision is best in the middle range of stimulus magnitudes, the ratio appearing not to hold if the stimulus is very weak or very strong. Some of the various ratios which have been worked out over the middle range of stimulus values are (125): lifted weights, $^1/_{20}$ to $^1/_{100}$; brightness of lights, $^1/_{60}$ to $^1/_{200}$; intensity of tones, $^1/_5$ to $^1/_8$; pressures on the skin, $^1/_{10}$ to $^1/_{30}$. For any given *S* the ratio may remain fairly constant while the inter-*S* variability may be great.

Weber's law as such allows one to make no assumptions concerning the increased variability of response as the magnitude of

the standard stimulus goes from low to high. However, if we find that the variable error increases as the size of a standard increases along a given dimension, we can deduce Weber's law from this fact. Increased variability means less precision of response, suggesting that a greater change would be necessary to bring about a clear-cut discriminable difference. A confirmation of this relationship is shown in the data on line drawing in Table 2 and Fig. 9.

Since Weber's law has been shown to be inaccurate in some instances, there have been several substitute laws attempting to state the relationship more precisely. We shall not consider them here. The general principle of the law is nearly universally true; the precise mathematical relationships may vary with different stimulus dimensions. Most of the recent work in sensory discrimination has not been concerned with whether or not Weber's or any other law holds. Data are usually plotted, however, so that a visual test of Weber's law can be made. Thus, the increase in the stimulus necessary to produce a difference threshold over the standard is plotted along the ordinate with the standard stimulus values along the abscissa. If Weber's law holds, the resulting graph should be a straight horizontal line.

Illustrative Studies Using the Method of Limits

A study of cutaneous sensitivity. Weitz (402) has used the method of limits to study the relationship between skin temperature and cutaneous (skin) sensitivity. In this problem *E* was testing a theory that sensitivity to pain is a function of a chemical reaction which is set off when a stimulus impinges on the receptor. Since chemical changes may be heightened by increased temperature, a deduction from the theory would state that raising skin temperature would increase the sensitivity to pain (lower the pain threshold). Pain was administered by electric shock from an inductorium.

The procedure consisted of placing the subject's arm in an arm rest, which had above it, tunnel-fashion, two metal shields with a space between. The shields were used in order to delimit the heated area. When the subject obtained a comfortable position the neutral electrode was placed in the subject's hand and taped there. However, in several trials the subject merely held this electrode and it should be mentioned that as far as the

results of this work are concerned there was no differential effect when the electrode was held or taped to the hand. The stimulating electrode was mounted in a wooden rod $^{3}/_{8}$ inch in diam. and 6 inches in length and was held by the experimenter on the dorsal surface of the forearm. When a "spot" was obtained which gave a clear "pain" or "pricking" sensation to electrical stimulation without any pressure or kinesthetic tug component, the electrode was kept on the "spot" and a normal intensity threshold was taken by the method of minimal change. As soon as this determination had been made the temperature of the skin, adjacent to the electrode, was taken. The electrode was never removed from the "spot" for the duration of the experimental session on that "spot." The radiant heater was then turned on and successive threshold measurements were taken using only ascending series. After each threshold determination, skin temperature was measured as before. This procedure was continued until the subject reported that the heat was uncomfortable or fatigue had set in due to holding a relatively fixed position. (402, p. 427)

Note that *E* actually used a modified method of limits for this procedure in that only ascending series of stimulations were used to determine the thresholds. This method is almost mandatory in this particular problem, since a descending series would mean that the shock was initially quite strong and was then gradually reduced in intensity.

Weitz used 6 *S*s in this experiment. It has been customary in experimenting on the sensory processes to use a few *S*s and take a great many rather precise observations on each rather than use many *S*s with a few observations on each. It is not unusual to report relationships derived from a single *S*. The results for one *S* in Weitz's experiment are shown in Fig. 11. The minimum strength of shock required to elicit pain response under normal temperature conditions is taken as 100 per cent. Deviations below 100 per cent indicate a lowering of the threshold (greater sensitivity) and deviations above would indicate decreased sensitivity. The skin temperature (base line) is indicated in terms of centrigrade degrees above normal, with zero degrees centigrade as the normal reference point. It is to be observed that the theory, i.e., that as skin temperature increases sensitivity to pain likewise increases, is confirmed up to a certain point, but beyond that point just the opposite relationship obtains. Weitz's theory will have to be revised to account for this finding.

We are able to appraise many of the environmental changes through our cutaneous sensibilities. Such sensations as heat, warmth, cold, surface pain, and light pressure (touch) are all mediated through organs lying near the surface of the body. There are wide differences in the sensitivity of the skin on various portions of the body. An experiment using the method of limits to demonstrate this point is easily performed in the laboratory. The problem is one of determining the two-point threshold of touch

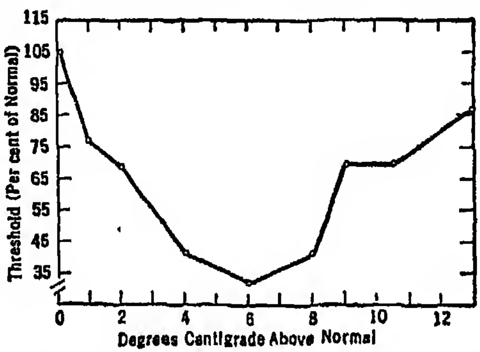


FIG. 11. The relationship between the changes in skin temperature and threshold for pain. The threshold for pain is taken as 100 per cent at normal temperature which is indicated on the abscissa as zero degrees centigrade. The data were obtained on one *S*. Adapted from Weitz (402).

on various body surfaces. In determining the two-point threshold *E* stimulates two points of the skin simultaneously and, by increasing and decreasing the distance between the points, determines what minimal distance must separate two points to give an impression of two stimuli. The device commonly used to stimulate the skin for determining two-point thresholds is called an *aesthesiometer*. The device may take many forms. A micrometer may be used; a compass can be fitted with two blunt ends to serve satisfactorily; even points of cardboard have been used. The main idea is to stimulate lightly and simultaneously two points on the skin for a short period of time, wish some means of measuring the distance between the two points stimulated. From a standard aesthesiometer this can be read directly from the scale. In determining the threshold for any one area of the body a straightforward method of limits may be used. *E* will start first with the

two points close enough together so that *S* reports only one point. The distance is gradually increased until *S* reports two points. In the descending series the points will be separated far enough so that *S* will easily discriminate two points and the distance will be gradually reduced. The raw data sheets look very much like the one shown in Table 4. The differences in the two-point threshold for different parts of the body are startling. The lips are extremely sensitive as are the finger tips, whereas the surface of the back yields a very large threshold value.

A threshold of form perception. One of the most fundamental discriminal processes is that of form perception. We distinguish roundness, squareness, triangularity, jaggedness, and an almost infinite number of variations around those modal descriptive terms. What sort of stimuli does it take to evoke these various interpretations? Does the response dimension always correspond to the physical dimension? What sort of errors are we likely to make in form perception? These are some of the questions which have been asked concerning the way in which we interpret the form or shape of things we "sense" in the environment. Many concepts have evolved to account for the observed facts. One of these concepts is *closure*. In a general sense this term indicates an observed tendency to perceive forms as being complete which are actually incomplete physically. If the outline of a circle is made up of discrete dots lying fairly close together we do not see a myriad of discrete dots. We would report first that we see a circle and if questioned further we would probably report that the circle is made up of a series of dots. We tend to fill in between the dots—we tend to close the unfilled gaps.

Bobbitt (30) has demonstrated experimentally that there is a definite threshold at which closure takes place in the perception of triangularity. To determine the threshold of closure for triangles, incomplete triangles were presented, as illustrated in Fig. 12. Figure 12*a* would probably not be seen as a triangle but rather as two discrete angles. Figure 12*b*, on the other hand, would probably be seen as a triangle—closure would take place. These two figures represent the extremes of the steps along the dimension defined in terms of the amount of the perimeter of the triangle present. A series of these finely graded forms was presented in ascending and descending order, each figure being exposed $^1/_{10}$

second. This short exposure period is accomplished by a device called a *tachistoscope*. *S*s were fully informed as to the nature of the experiment. In responding to the figures *S* was to report whether he saw "twoness" (two angles) or "oneness" (triangle). From this procedure, thresholds of closure of considerable stability were derived. Depending on the size and shape of the "triangle," approximately 68 to 72 per cent of the perimeter had to be present before the figure was seen as a triangle. There were no systematic changes in the threshold throughout the experiment, i.e., practice did not seem to influence the threshold and fatigue was minimized by adequate rest.

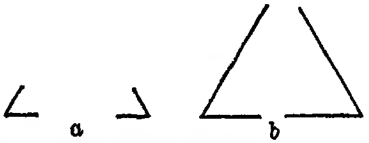


FIG. 12. Illustration of two extreme figures used by Bobbitt (30) in determining the closure threshold for triangularity. Figure *b* is likely to be seen as a triangle, whereas this is not usually true of Fig. *a*.

Auditory thresholds: Are two ears better than one? As we have said, the method of limits is used primarily to determine sensory thresholds. This experiment will consider some of the problems involved in determining auditory thresholds. For any given frequency of the sound wave the lower absolute threshold of intensity may be plotted. The physical intensity of a tone is measured by a scale, the units of which are called *decibels*.

Monaural intensity thresholds may be determined by feeding the sound into one ear and by the use of heavily cushioned earphones which prevent bone conduction of the sound wave to the other ear. An experiment by Shaw, Newman, and Hirsh (350) was designed to determine whether or not a binaural threshold was lower than the monaural threshold. The experiment demonstrates the method of limits in considerably modified form, the modification being necessary for this particular problem.

The tones used in this experiment were produced in pure form by commercially built oscillators. *E*s presented the tones in "beep-beep-beep" fashion rather than continuously. It is necessary, in designing the experiment, to control for slight differences in the sensitivity of the earphones. Shaw, Newman, and Hirsh indicate

that to get the best estimate of the thresholds for binaural and monaural presentation the following six conditions are necessary:

Right ear alone: Earphone *A*
Right ear alone: Earphone *B*
Left ear alone: Earphone *A*
Left ear alone: Earphone *B*
Both ears: Earphone *A* on left, *B* on right
Both ears: Earphone *A* on right, *B* on left

For each of these six conditions a threshold would be determined. The mean of the first two conditions gives the best estimate of the threshold for the right ear, whereas the mean of the third and fourth conditions yields the estimate for the left ear. The mean of all the first four conditions gives the best estimate of monaural thresholds. The mean of the last two conditions provides the estimate of the threshold for binaural hearing. If there are differences in the sensitivity of the earphones, the series of conditions spreads the influence of these differences equally over the binaural and monaural measurements. Furthermore, since there may be practice effects, all *S*s should not take the conditions in the order listed. In this particular experiment a random order of the six conditions was determined for each *S* by drawing numbers from a box. All *S*s went through all six conditions but in different order.

For careful determination of thresholds, well-practiced *S*s are desirable. *S* usually works in a sound-proof room. When the tone coming through the earphones is of very low intensity, noises emanating from the body and the head of *S* appear to assume considerable magnitude so that *S* must be skilled in distinguishing these noises from the tone in the earphone. The modified method of limits used in this present experiment required *S* to determine his own threshold. The tone was initially set considerably above threshold. *S* then manipulated a dial thereby reducing the tone to near threshold level. At this point *S* would "play around," first increasing and then decreasing slightly the intensity of the tone until he believed that he had the best estimate of his threshold. One of the characteristics of the tone at the threshold is that it disappears and reappears even without changes in the physically measured intensity being present. This again emphasizes the non-static character of thresholds. Results from this experiment for

14 different frequencies show that the binaural threshold is one to two decibels lower than the monaural threshold. This difference seemed to be independent of frequency and was also found when speech sounds were used instead of pure tones.

Measuring animal thresholds. The illustrative experiment we will consider here concerns the measurement of differential thresholds in the cat. Such a study poses some interesting methodological problems. How is the cat going to tell *E* when it notices a difference in the intensity of a tone? Verbal communication being out of the question, some other method must be devised. Basically the method consists in first training the animal to make a response to a change in stimulation. To do this, the conditioning technique, to be discussed fully in a later chapter, is used. In this particular experiment by Rosenzweig (322), the cat was placed in a rotary cage the floor bars of which were wired so that the cat could be given an electric shock. In setting up the conditioned response a tone of a particular frequency was sounded and the intensity of that tone increased noticeably (to the human ear) by a series of steps. After some change in intensity had taken place the cat was shocked. The cat soon learns that a "change in intensity means that I am going to get shocked." After several trials, then, the cat starts to run whenever a change in intensity is noted. If the cat started to run soon enough no shock was given. An essential control to this procedure is that the sound be kept on at all times so that the cat does not learn to respond to the onset of a tone rather than to changes in its intensity.

Once the animal has been trained to respond to a change in intensity, the measurement of the threshold can proceed. The learning has established a communication system between the cat and *E*. *E* can gradually increase (or decrease) the intensity until the cat responds by running. When the cat responds in this fashion it is assumed that the animal has discriminated a difference in intensity. *E* has thus measured the stimulus change necessary to produce a just noticeable difference in the animal. In this experiment it was found that the difference threshold in the cat was slightly larger than difference thresholds in humans at corresponding frequencies.

The above are enough samples of the uses to which the method of limits can be put. We cannot compile the mass of facts for all

the various stimulus dimensions of the various sense modalities, but we should understand the variety of threshold measurements for which the method is adaptable. We have not considered much in the way of theory which prompts most investigations. To a large extent the theories relate to physiological structure of the sense organs and the relationship of these to the central nervous system. In order to present the theories in any meaningful fashion we would have to expand the physiological discussion far beyond the scope of this book.

The method of limits is by no means restricted to the study of laboratory problems on "pure" sensation. The method may, if one desires, be applied to a problem which has immediate practical importance. For example, one investigator (16, 17) has recently published a series of experiments in which he has determined the optimum form, size, and light-dark color combination of numbers for automobile license plates. His determinations were made entirely by the method of limits under conditions of both day and night driving. As a consequence of the work, he has been able to patent the system of license plates which gives optimal recognition thresholds. The practical implications are obvious; the policeman can see license plates at a greater distance than heretofore even though the plates themselves are no larger.

Summary of Section on Method of Limits

1. The method of limits has been used largely to determine lower absolute thresholds and difference thresholds. Essentially the method consists in varying a stimulus along a given dimension and finding the point at which it no longer evokes a response (absolute threshold) or at which point it is just discriminably different from a standard stimulus (difference threshold).

2. Constant errors of habituation and expectation were discussed. The inevitable variable error is present as in most psychological measurements. The variability of the threshold measurements emphasizes the non-static character of the threshold.

3. Relationships between the size of the difference threshold and the standard stimulus were discussed. The simplest statement of the relationship is known as *Weber's law*, which states that the ratio between a standard stimulus and a difference threshold for a given stimulus dimension is a constant fraction of the standard

stimulus. The law holds roughly over a wide range of magnitudes.

4. Samples of experimental procedures and results have been presented as typical of the problems to which the method of limits is applicable.

CHAPTER III

Discriminal Processes: Methods of Study II

We shall continue the discussion and illustration of the methods which have been used to study the discriminal processes. In this chapter we shall consider two methods: (1) the method of constant stimuli; and (2) the method of equal appearing intervals.

THE METHOD OF CONSTANT STIMULI

The Method

This method is one of the oldest psychophysical methods. Traditionally, it has been used for much the same purpose as the method of limits—to measure thresholds. One variation of the method has been used to measure absolute thresholds, another to measure difference thresholds. As we have done with the other methods, we shall present first an hypothetical pure case of procedure and results.

The method applied to absolute thresholds. In the method of limits, *S* is presented a stimulus of *gradually changing* magnitude and is asked to report when the experience ceases (descending series) or when the experience starts (ascending series). *In the method of constant stimuli each trial consists of the presentation of an invariable stimulus with S asked to report its presence or absence.* On successive trials, stimuli of different magnitudes are presented, some of them being sufficiently great to be above threshold, hence, experienced, and some weak enough to be below the threshold so they are seldom experienced. These stimuli are *not* presented in an ascending or descending order of magnitude but rather in a random order. Let us illustrate the method by working with a lower absolute visual threshold.

E, by preliminary work, determines the approximate value of

S's absolute threshold. Then a series of stimuli is chosen extending from well below to well above this threshold. *E* will choose 4 to 10 stimuli to make up the series. The weakest stimulus will be weak enough so that rarely will *S* report its presence; the strongest stimulus will be strong enough so that *S* will almost always report its presence. The other stimuli form equal steps on a physical scale between these two extremes. *E* determines a random order for presenting these stimuli. Consequently, one of weak magnitude may be presented first; a strong one next, and so on. Each stimulus is presented an equal number of times, perhaps a hundred times or more.

Let us say we have chosen six light intensities with one unit of a physical scale separating each successive magnitude. The lowest stimulus magnitude has a value of 7, the highest a value of 12. These are presented to *S* in random order 100 times each, with instructions to report at each trial if a stimulus is "present" or "not present." As in the method of limits, the experiment would be carried out in a dark room, with *S* well habituated to the situation and watching a fixation point so that he knows where to expect the stimulus. After the presentation of each stimulus *E* records whether *S* responded positively or negatively. For each stimulus value *E* would obtain a percentage or proportion, indicating the relative frequency with which *S* reported the stimulus present. In this hypothetical experiment the following data were obtained:

Stimulus Value:	*7*	*8*	*9*	*10*	*11*	*12*
Percentage of times reported present:	3	11	35	68	87	99

The absolute threshold is that lowest stimulus value which is noticeable 50 per cent of the time. In the data above we note that a stimulus value of 9 was perceived 35 per cent of the time; a stimulus value of 10, 68 per cent of the time. The threshold must lie somewhere between these two values since, presumably, some value between 9 and 10 would, if actually presented a number of times, be judged present half the time and absent half the time.

From the above data there are two relatively simple ways in which a close approximation of the threshold can be obtained. One, a graphical method, consists in plotting the data with the stimulus values along the abscissa and the percentage values along the ordinate. Drawing a smooth curve through, or as near as

possible to the empirical points, will allow us to discover where the curve passes through the 50 per cent point. Such a curve is shown in Fig. 13. Dropping a vertical to the abscissa from the 50 per cent point we find the stimulus value is about 9.4. This value will vary slightly, of course, depending on the precision with which the curve is drawn, but it represents a fair approximation of the absolute threshold.

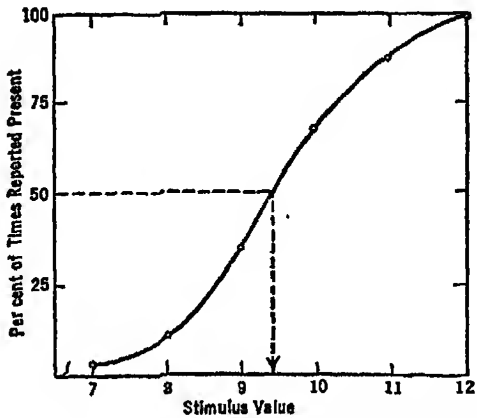


FIG. 13. Graphical determination of the absolute threshold. The threshold value is approximately 9.4.

A second, and more accurate method, uses the same technique as employed in finding the median of a distribution of scores. The total "percentage distance" between the stimulus value 9 and the stimulus value 10 is 33—obtained by subtracting 35 from 68. We know that the 50 per cent point (defining the threshold) is 15 percentage points above 35 per cent, or 15/33 of the stimulus difference between 9 and 10. The step interval being 1, the "distance" we have to go into this interval is 15/33 by 1, which is .46. Therefore, the stimulus value which should be judged present 50 per cent of the time would be 9 plus .46, or 9.46, a slightly larger figure than obtained by the graphical method. This method of finding the threshold is sometimes called *linear interpolation.*

Determination of thresholds by the method of constant stimuli has one distinct advantage over threshold determination by the method of limits. In the method of constant stimuli the randomization

of the stimuli probably eliminates the errors of expectation and habituation which may mar the threshold measurements by the method of limits. In fact, all things considered, there is reason to believe that the method of constant stimuli yields a somewhat more precise threshold than does the method of limits (125). Examination of current psychological literature, however, shows that the method of limits is being used much more widely for this purpose.

The method applied to difference thresholds. The determination of difference thresholds by the method of constant stimuli requires *a standard stimulus against which other stimuli of varying magnitude are judged.* Let us take for our illustration of the mechanics of the procedure a discriminal process which is a classic laboratory exercise; namely, judging weight differences.

For this illustration the standard weight against which all other weights will be judged is 100 grams. We construct a series of comparison weights which varies from the standard only in weight and not in any other characteristic, i.e., we make these weights the same size, same texture, and so forth. Small aluminum cans of identical size have often been used with lead shot determining the weight. The shot, regardless of the amount in the can, is held in fixed position by the addition of cotton. *S*, blindfolded while making the judgments, can get no cues for judging differences other than those supplied by the independent variable—the weight. In choosing the comparison weights it will be necessary to make the extremes of the series clearly different from the standard, i.e., the lightest in the series should be judged lighter than the standard nearly 100 per cent of the time, and the heaviest, judged heavier than the standard almost 100 per cent of the time. In our series the lightest weight is 90 grams and the heaviest, 110, with step intervals of two grams so that the comparison series is as follows: 90, 92, 94, 96, 98, 102, 104, 106, 108, 110. Each variable is compared with the standard many times. As in the determination of the absolute threshold, comparison weights will be presented in random order.

The two stimuli (standard and variable) are presented to *S in succession*, *S* "hefting" them one at a time with one hand—he does not hold them simultaneously, one in each hand. If the standard were always lifted first and the comparison weight second, a

constant error might be introduced from the order of presentation as a result of the time interval between stimuli. Consequently, in half the cases the standard is lifted first and the variable second; in the other half the reverse order holds. *S* must judge whether the second weight lifted is "heavier" or "lighter" than the first. *E* determines the percentage of times each comparison weight is judged heavier than the standard.

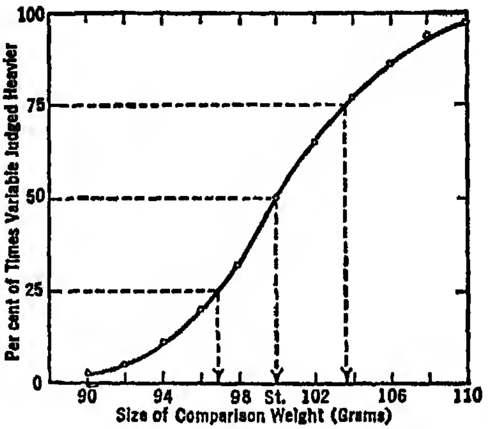


FIG. 14. Representation of the results of the weight-lifting experiment using the method of constant stimuli. A standard weight of 100 gm. was used and compared with each variable weight. The points for the curve are determined by the percentage of times each weight was judged "heavier than" the standard. As depicted here, the point of subjective equality is 100 gm., the lower difference threshold slightly over 3 gm., and the upper difference threshold slightly less than 4 gm.

Figure 14 is a graphical representation of the results obtained in this hypothetical experiment on a single *S*. Since we have counterbalanced the order of presentation of the standard and variable weights, no constant errors due to procedure should be present. Hence, the point at which our curve crosses the 50 per cent line should be directly above the standard weight of 100 grams as indicated on the abscissa. The value will probably vary slightly from the value of the standard due to variable error. The value can be determined also by finding the median value between the proportion for the weights immediately below .50 and that immediately above. This we have done in the previous illustration of

the absolute threshold. This value, it will be realized, is the point of subjective equality—the weight which is apparently equal to the standard weight.

The procedure for finding the difference threshold from data obtained by methods similar to the present is a point on which there has been disagreement (125, 422). For our purposes a simple, but slightly rough way may be used. This method meets all the requirements of the statistical difinition of a difference threshold and becomes very similar to the method employed in the method of limits. Here, as in the data obtained by the method of limits, we have two difference thresholds, one above the point of subjective equality and one below the point of subjective equality. The upper difference threshold would be that stimulus range from the point of subjective equality to the upper point at which the judgments are correct 50 per cent above chance. In terms of the graph this would actually be at the 75 per cent level. The lower difference threshold would be the stimulus range between the point of subjective equality and the lower stimulus value which is correct 50 per cent oftener than chance. This is the 25 per cent point on the graph.

This determination of difference thresholds may seem a little confusing at first, but if you will re-check on the definition of a difference threshold and study the graph, it should become clear. Note, for example, how the curve would look if we assumed a very small difference threshold—much smaller than shown on the graph. In such a case *S* might get all or nearly all correct responses when comparing the 96-gram weight with the 100-gram standard and the 104-gram weight with the standard. This would mean that the new curve would rise at a much sharper angle. Then if verticals were dropped from the point of subjective equality and the 25 per cent and 75 per cent points on this hypothetical curve, the stimulus ranges defining difference thresholds would be smaller than those shown in Fig. 14. On the other hand, if the curve is flattened a great deal and a greater range of comparison stimuli used, it can be seen that the stimulus difference which is noticeable 50 per cent of the time would be greatly increased.

The graphical determination of the 50 per cent point above chance is probably as accurate as when the median formula is used, since the curve is less linear at these points (25 per cent and 75

per cent) than in the middle range. Hence, some error is involved in assuming linear interpolation. If the curve drawn does not deviate widely from any of the empirical points, and if the empirical points are based on many judgments, the graphical determination of the difference threshold as suggested here is accurate enough for elementary laboratory work; hence, other methods will not be duscussed.

Supposing *S* lifts two weights and can distinguish no difference? In the procedure we have been discussing, *S* could respond only with "heavier than" or "lighter than," even when he could perceive no difference between the weights. Some *E*s have allowed *S* to give an "equal" response although there is apparently not a great deal gained by it (125, 422). Furthermore, it has been shown (42) that even when *S* adamantly asserted that he could not distinguish a difference in two weights his forced choices were better than chance, i.e., his decision as to which was heavier concurred with the physical measurements more often than chance would allow. *S* believed he made sheer guesses and yet these guesses were more often correct than incorrect.

As in all the methods thus far discussed, there are deviations from the modal procedure which defines the present method. One of the variations on the method of constant stimuli has been called the *method of single stimuli.* In this method *S* does not have a standard for comparative purposes but instead makes an absolute judgment. Rather than being asked to judge which of two weights is heavier, he is asked to make a judgment as to the actual weight of the stimulus. It is presumed that a subjective scale or subjective standard of some sort is built up and this subjective standard becomes the comparison weight. The weight guesser at the county fair uses such a method. He probably has a series of subjective standards built up through experience so that each new contributor is judged relative to one of these standards. The standard by which the new individual is judged is determined by an estimate of the height, the bone structure, the "width," solidness, and so forth, of this individual. If the guesser's standards are fairly well fixed he can allow himself a few pounds margin of error and do rather well.

We are going to consider the method of single stimuli more fully in the next chapter since most of the current work with

this method concerns stimuli which are not described by physical dimensions.

Illustrative Experiments Using the Method of Constant Stimuli

One of the most widespread uses of the method of constant stimuli has been in investigating a constant error called the *time error*. Because a considerable amount of work is being done currently on this phenomenon, we shall consider several experiments specifically designed to study it. These reports, as we shall see, provide a good working illustration of the relationship between experimentation and theory in a restricted area of behavior.

The time error. When judgments are made concerning the relative amount of a given characteristic of two objects, and when these objects are presented successively, there is the possibility that processes going on during the time interval between the two observations may influence the judgment. Thus, if *S* lifts two 100 gram weights in succession, it is possible that the "lifts," separated in time, may affect his observations. Obviously, if this time interval does influence the judgments in a systematic fashion we have evidence for a constant error, the *time error*. In the weight-lifting experiment such processes usually produce a constant error in the direction of judging the second weight heavier than the first, even though objectively they are the same in weight. To determine if, and in what direction the error is operating, *E* subtracts the standard stimulus value from the stimulus value defining the point of subjective equality. In the weight-lifting experiments we usually have a *negative time error;* the second of two equal weights is judged heavier than the first. If the second of two equal stimulus magnitudes is judged less than the first, we would have a *positive time error.*

In a well-controlled weight-lifting experiment rather elaborate rotary trays have been devised to present the stimuli to *S*. The weight to be lifted is presented directly under the hand of the blindfolded *S* so that all he has to do is pick it up for a second or two, get an "impression" of its weight, and then replace it. *E* immediately presents the second stimulus (by rotating the tray) so that it too is lifted merely by lowering the hand and grasping the container. After the second weight is lifted and *S* responds by indicating that the second weight is either heavier or lighter than

the first, *E* moves the tray on to another stimulus, and so on. Even with relatively short intervals between "lifts" of the two weights, perhaps 2 or 3 seconds, a negative time error may be operating so that the second will be judged heavier than the first more often than expected by chance. Of course, the over-all influence of the time error will be cancelled if the variable weight is different from the standard, and if these weights are presented in a series of judgments so that half the time the standard appears first and half the time the variable appears first. The data would then be fractionated to determine the extent of the time error.

If the time error itself is being studied, the counterbalancing procedure is not used. Rather the standard is always presented first with the variable stimuli second. Then, if we plot the curve relating the percentage judged greater to the magnitude of the variable stimuli (as in Fig. 14), the point of subjective equality may be used as a means of determining the extent and direction of the time error. If the point of subjective equality is a value greater than the standard, the time error is positive; if the point is less than the standard, the time error is negative.

Illustration of the time error. As suggested by the name of the time error, the length of the time interval between the two stimuli is an important variable in determining the extent of the error. The relationship is exemplified in a study by Postman (297). The standard stimulus was a 1000 cycle tone, always set at an intensity of 75 decibels above the threshold. Comparison tones varying in intensity both above and below the standard were used. The standard was always presented first, the variable second. The effect of four different time intervals between the standard and comparison tones was determined, the four intervals being 1, 2, 4, and 6 seconds. Four *S*s took part in the experiment and, as shown in Fig. 15, the time error becomes greater and greater as the interval between the standard and comparison tones increases. At first the error is slightly positive but becomes negative and increasingly so as the interval becomes longer.

Time error theory. In order to understand the implications of some recent experiments, we must have a general comprehension of the nature of the theories which have been formulated to account for the time error, specifically, the negative error, since such an error is usually found. Most prominent in terms of the number

of adherents and the number of publications built around it is the *trace theory*. In general terms the trace theory postulates that the first of the two impressions fades or weakens between its presentation and the presentation of the second stimulus. If this is true, the second stimulus would be judged greater than the first because it is compared with a faded impression of the first. Thus, if the two stimuli were objectively equal, the second would be

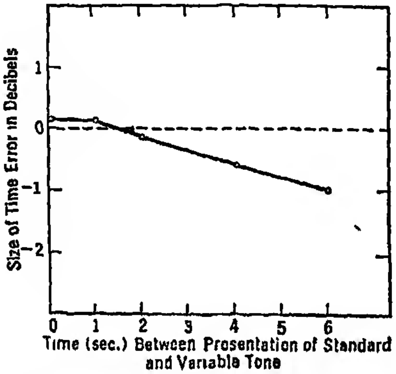


FIG. 15. The time error as a function of the length of the interval between the standard and the variable. The standard and comparison tones were 1000 cycles. The standard was set at 75 decibels intensity and the comparison tones varied above and below that level. A positive time error indicates that the point of subjective equality was greater than the standard; a negative error indicates the point of subjective equality was less than the standard. Data from Postman (297).

judged heavier than the first and give a negative time error. For our purposes we need not—as indeed we cannot—specify the exact nature of the trace of the stimulus other than its basic characteristic of fading with the passage of time.

Recent studies have shown that in some instances where negative time errors have been found, a slight change of conditions may produce a positive or zero time error. These findings must necessarily have a bearing on the theory suggested to account for the errors. We shall consider some of these studies and their implications for the theory.

Factors producing positive and negative time errors. An experiment by Marchetti (245) was concerned with the judgment of visual extent. Instructions to the *S*s were as follows:

> You will be shown a black line on a white cardboard. After a brief interval another line will be shown to you. You are asked to judge the length of the second line in terms of the length of the first, i.e., whether the second is *equal* to, or *shorter* or *longer* than the first. Make your judgment as soon as you can after you see the second line (245, p. 258).

In all presentations the standard line was shown before the variable line. It will be noted that Marchetti allowed "equal" judgments. To determine the point of subjective equality from such data the median of a distribution of equal judgments is employed. *E* used several different line lengths and several different intervals between the presentation of the standard and the comparison line, but in most cases there was a clear-cut indication of a *positive* time error; that is, the point or length of subjective equality was greater than the standard length. Under certain conditions, however, a negative time error was produced. If the two lines to be compared were of equal length but with one drawn on a smaller card than the other, the one on the smaller card appeared longer. Hence, placing the second of the two stimuli on the smaller card produced a negative time error. In addition to the length of line as such it became evident that other factors in the situation entered into the judgments. The relative length of the line to the card on which it was drawn determined, to a certain extent, the impression *S* carried over to the judgment of the second stimulus. This is an illustration of a process called *assimilation.*

Marchetti used tachistoscopic exposure of the stimuli in his experiments. McClelland (226), in a series of closely coördinated experiments designed to study the time error in the judgment of visual extents, used screen exposures in which the end of the brief exposure period was brought about merely by turning off the projector lamp. Under such conditions McClelland found a negative time error—in direct contradiction, it appeared, to Marchetti's results. McClelland observed, however, that in turning off the projector lamp at the end of the exposure period the stimulus on the screen did not cut off abruptly but gradually faded as the filament in the lamp cooled. This suggested to him that the negative time error which he had found might be due, not to a faded trace of the first impression, but to a fading in the stimulus itself so that the final impression of the first stimulus would be one of "lessness" when compared with the second of objectively

equal length. To test this hypothesis he changed the procedure to make it comparable to Marchetti's: namely, so that (as with a tachistoscope) the beam of light from the projector would be cut off quickly; hence the line on the screen would not fade. Under such conditions the time error appeared as positive instead of negative and thus confirmed Marchetti's findings.

McClelland's results led him to survey the conditions under which the time error had been found in other sensory modalities, and his conclusion is that in many cases the time error may be due to a "fading in the sensation itself, caused either by a fading stimulus source or by receptor-adjustments following stimulation" (226, p. 94). Applying this hypothesis to lifted-weights data, for example, McClelland points out that the last impression obtained from the first of the two weights lifted comes when the weight is *put down*, when the muscles are no longer holding the full weight. The final impression of the first weight might be somewhat less than it was when lifted and held. To test this possibility a weight lifting experiment was performed in which the first weight was *dropped* instead of being set back down. Preliminary results from this experiment confirm the prediction that under such conditions the negative time error should disappear. McClelland's basic conclusion is that some instances of negative time error can be explained on the basis of artifacts in the experimental situation and that a trace theory need not be used.

But what about the positive time errors in the judgment of visual extents? Both Marchetti's and McClelland's work suggest that the concept of assimilation is adequate to account for positive time errors. In one of McClelland's experiments *S*s made judgments of visual extents in a room which was completely darkened so that no contours of the screen or room were seen. Under such conditions the positive time error almost disappeared. It is presumed that the removal of the room contours by imposing darkness left nothing with which the stimulus could assimilate. What does appear obvious is that the time error is subject to considerable variation by slight changes in the stimulus situation. *E*s must state in detail under what conditions the positive, negative, or zero time errors are found.

So much for the time error. We have seen that the time error can be changed markedly by varying conditions. We have also

seen that theory has played a large part in directing the experiments which have been performed in recent years.

Judgment time and confidence. It is a common observation that certain decisions require longer to make than do others. When decisions take a long time it appears to be caused by the fact that the differences between the alternatives are slight. In simple form, if we were to judge which of two weights was heavier, it would take less time to judge an 80 gram weight against a 100 gram weight than it would a 96 gram weight against the 100 gram. The greater the similarity in amount of a physical attribute possessed by two stimuli, the longer it will take to make a judgment concerning the relative amounts of the attribute.

Studies by Festinger (94, 95) show that there are other response characteristics associated with judgment time, one of these being the confidence in the judgment once it is given. In studying this problem, Festinger used the method of constant stimuli but presented the two stimuli simultaneously so that the time error did not enter into the results.

Two black metal boxes were used for presentation of the stimuli. Ordinary electric light bulbs were placed inside each box. In front of each box was an opening, $^3/_8$ inch wide, extending the length of the box. A metal strip, placed through grooves on each side of the opening, could be moved back and forth so as to cover up all or any desired portion of the open part. The opening itself was covered with heavy ground glass through which the light could shine.

The boxes were mounted upon a table 31 inches high so that the openings in front of the boxes were vertical. The distance between the neighboring edges of the two vertical lines was 9 inches. A black cardboard screen with two slits in it was mounted in front of the apparatus so that the only thing visible from the front was an homogeneous black background and the two vertical lines. The experiments were always carried out in total darkness so that nothing could be seen until the two vertical lines were illuminated.

The *S*s sat in a chair 130 inches away from the apparatus and midway between the two vertical stimuli.

The two lines were presented simultaneously to the *S* for 0.5 second. The duration of the presentation of stimuli was controlled by an electron tube tachistoscopic timer. When the tachistoscope was turned on, thus illuminating the vertical lines, an electric stop clock started. After 0.5 second the lights went off automatically, but the clock kept turning until the *S*'s response activated a voice key which stopped the clock. . . . After

each response, the experimenter, using a very dim flashlight, recorded the proper data and set the apparatus for the next presentation.

The stimulus line at the left of the *S* was always kept at a length of $7^1/_2$ inches. This we shall refer to as the standard. The length of the other stimulus line was adjusted by means of the sliding metal strip. . . . This line we shall call the variable stimulus.

Fifteen positions of the variable stimulus were used, seven shorter than the standard, seven positions longer than the standard, and one position equal to the standard (94, pp. 291-292).

In determining the relationship between judgment time and the confidence in the judgment, the following special instructions were given:

After making each judgment you are to express your confidence that the judgment was correct in the following terms: 50 per cent confidence means average or median confidence, the amount of confidence one usually has in most of one's judgments; 75 per cent means confidence definitely above average, the amount of confidence you have when you are quite sure you are right; similarly, 25 per cent means definitely below average confidence, the amount you have when you are only guessing; 100 per cent confidence is at the extreme end of the scale of confidence and means complete certainty, no possibility of error; similarly, zero confidence is at the other end of the scale of confidence and means no confidence in your judgment whatever; you are just as likely to be wrong as right.

Thus, you are to construct a "mental scale" of confidence with 50 per cent in the middle, 25 and 75 per cent on each side of the middle range, and 100 per cent and zero per cent way out at the ends of the scale (94, p. 293).

*S*s were given preliminary practice until they were well acquainted with the procedure. On the experimental trials, each *S* made 40 judgments for each of the 15 variable lengths. The results of one *S* are indicated in Fig. 16, showing the relationship between the difference in the actual lengths of the standard and variable and the confidence expressed in the judgment. The confidence scale along the ordinate is constructed so that the zero point, the value to be used when *S* has no confidence, is in the middle, with a scale to 100 running above and below this. The relationship is quite clear: namely, that as the difference in the lengths of the variable and standard increase, confidence in the judgment likewise increases. Results of other *S*s, not plotted, showed similar results.

Related work. The rationale of the method of constant stimuli underlies work in other areas. The revised Stanford-Binet Test of Intelligence (376) is so constructed that the scoring procedure follows the basic philosophy of the constant methods. *S* is given first a series of problems all of which he can answer correctly. This is done by presenting problems fitted for an age level below that of *S*. The examination proceeds by increasing the difficulty

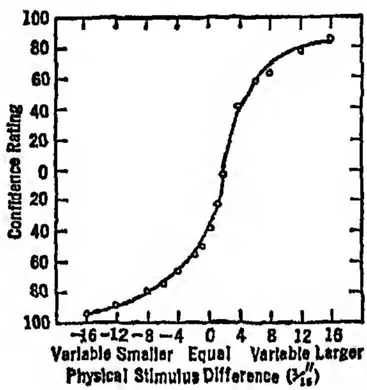


FIG. 16. Relationship between physical stimulus differences of two lines and the confidence expressed in judging these differences. *S* was asked to judge whether one line was shorter or longer than the other and then express the confidence rating. The curve shows that the greater the difference in the length of the two lines the greater the confidence in the judgment. 100 indicates high confidence; 0, low confidence. Adapted from Festinger (94).

of the items in a series of steps until all items at a given level are missed. *E* has explored the range of responses from the point where all items are answered correctly to the point where all items are answered incorrectly. *S*'s mental age is certainly not at the highest age level where all items were correctly answered and just as certainly not at the lowest age level where all were failed. The mental age must be somewhere in between. The highest level where *S* correctly answers all problems is called the *basal age*. Above this is a series of age steps for which the problems are increasingly difficult and *S* will miss more and more until the level is reached where none is correct. To the basal age is added the number of correct responses obtained at the various ages

above the basal age. The best guess as to the mental age is near the age level where *S* gets 50 per cent of the problems correct, though in actual practice it is somewhat higher. Since different sets of problems are related to different age levels, *S*'s mental age is an estimate of the point or age level associated with the set of problems on which he scores 50 per cent correct. Mental age is then used in deriving intelligence quotient (IQ).

The Seashore Measures of Musical Talents (342) use the method of constant stimuli in determining the musical aptitude of *S*. For example, in determining the acuity of pitch discrimination, *S* is presented with a series of pairs of tones and is asked to judge whether the second tone is higher or lower in pitch than the first. These stimuli are presented phonographically and the series includes pitch differences which are easily discriminable and differences which are not perceptible. If *S* on the average is able to discriminate only large differences as compared with the average, it indicates a large difference threshold and a likelihood of rather low musical aptitude. The smaller the difference threshold the greater the sensitivity to pitch differences and, when taken in conjunction with small thresholds along dimensions such as intensity and rhythm, the greater the probability of *S* being a successful musician.

Summary of Section on Method of Constant Stimuli

1. The method of constant stimuli has two major variations:

a. For finding absolute thresholds, single stimuli in the region of the threshold are presented and *S* notes their presence or absence. The threshold is determined by calculating the stimulus value which will elicit the response of "present" 50 per cent of the time.

b. For finding difference thresholds, *S* is presented pairs of stimuli and, after the presentation of each pair, judges whether the second stimulus is of greater or less magnitude than the first. In each pair is a stimulus which remains constant with regard to the amount of the stimulus characteristic under consideration. This is called the standard stimulus. The second member of the pair is the variable or comparison stimulus in that it is, in successive pairs, varied in amount extending both above and below the amount of the standard. Each pair is judged many times and the standard will follow and precede the variable an equal number of times. The variable stimuli are presented with the standard in random order, thus differentiating the method from the method of limits.

2. A major source of error arises when the two stimuli are presented successively rather than simultaneously. This constant error, the time error, has been widely investigated and much of the interest in the constant method seems to be perpetuated because the method allows for the study of this error. Several recent studies pertaining to the time error were reviewed.

THE METHOD OF EQUAL APPEARING INTERVALS

Most boys have had the experience of laying out a sandlot baseball diamond without benefit of a measuring tape. The procedure consists of finding a suitable place for home plate and then attempting to place the bases so that they are equally spaced and at distances that seem to approximate those of the official diamond. In laying out this diamond the idea is to make the distances between bases appear as equal intervals. Because we recognize the error in our instrumentally unaided judgments, we often have "stepped off" the distance to aid in equalizing the intervals. *It so happens that the visual equalization of distances is a well-*practiced discriminal process in our culture and we can be surprisingly accurate except when certain distractions are added to bias our judgments, as in the case of the Müller-Lyer illusion. The breaking up of a "sense distance" into equal psychological intervals is the major purpose for which the method of equal appearing intervals was devised.

Before considering the details of the method, we must give some attention to problems of measurement which thus far we have skirted. It is necessary to understand the implications of these problems in order to comprehend the purposes and uses of the method of equal appearing intervals.

Scales

Quantified dimensions which reflect characteristics of physical phenomena are often called *extensive scales.** The major characteristic

* Do not take the word *extensive* literally. As used in the present context it does not imply "large, or extended in space." Likewise, the term *intensive*, to be introduced shortly, is not to be given its usual meaning. Their special meanings as customarily applied to the subject matter we are discussing, will be made clear in the text.

of an extensive scale is that the unit of measurement is the same size throughout the entire dimension. To put this briefly, "an inch is an inch," or, "a foot is a foot"; the distance from 2 to 3 inches is exactly the same as the distance between 122 and 123 inches. We have seen that psychological response dimensions, even those which are largely a function of a single physical dimension, seldom show a one to one relationship with that physical dimension. Increasing the frequency of the sound wave from 100 to 200 cycles does not thereby "double" the psychological phenomenon of pitch. Since the psychological dimension does not directly parallel the physical dimension, we would know that a unit change in the physical stimulus at different places along its extensive range would not elicit equally differing psychological experiences. Weber's law implies these facts. It thus remains somewhat of a problem, and an intriguing one, to construct a *psychological* scale in which the units are experientially equal throughout the entire dimension.

If we were to construct a psychological scale with equal units, what unit of psychological experience might we use? One which has been suggested as being the smallest possible unit, having a counterpart in experience, is the difference threshold or just noticeable difference. Weber's law has given us the basic information that the amount of change in the physical stimulus necessary to bring about a just noticeable difference will vary as a function of the absolute magnitude of the physical stimulus. So, why not use the difference threshold, or some multiple thereof, as the unit of the scale and relate each difference threshold to a change in the physical dimension? For example, we might start near or at the lower absolute threshold for pitch, and add successive differential thresholds until the entire range is covered. Since each successive threshold is determined in exactly the same way regardless of the magnitude of the stimulus change, each should be experientially equal to all others.

Such a procedure has been carried out on certain dimensions. There is, however, disagreement as to whether all differential thresholds determined in this fashion are equal. For some dimensions, the subjective experience associated with the ninth difference threshold in the series is not equivalent to the subjective experience associated with the twenty-ninth, merely to pick two

numbers as illustrations. What this means is that *S*, confronted with just noticeable differences from two different parts of a series, will not necessarily judge these differences to be subjectively equal. If this is true, and it has been reported in some cases, our units have a rather superficial or false equality—a difference threshold is not a difference threshold in the same sense that an inch is an inch. We must conclude that in spite of the quite obvious equalizing of the difference thresholds by the method used to measure them individually in a given series, we cannot use them as units of a scale without first demonstrating their phenomenal equality. For some dimensions of psychological experience they seem to be equal, for others not.

In addition to the feature of equal units which we have discussed, extensive scales have an *absolute zero*. This is easily seen in the case of physical measurements. Zero length, zero weight, and so forth, can be shown to be the logical limit to the lower end of the scale. For the scaling of psychological sensory experiences, the lower absolute threshold is usually assumed to be the absolute zero.

Another property of extensive scales is that called *additivity*. This means that if we add two equal units together by the usual arithmetic operations, the resultant quantity will be twice as great as either alone. In the case of physical scales, such as an inch scale, this is easy to visualize. If we add the difference between 2 and 3 inches to the difference between 3 and 4 inches, the resultant will be twice as great as either difference alone. For psychological measurements the feature of additivity would mean that if we have two psychological units which are equal (such as two difference thresholds), and if we "add" them together, they should give an experience twice as great as either alone. We cannot attend to the problems involved in fulfilling this qualification for psychological dimensions except to say that the evidence is inconclusive as to whether or not psychological scales have ever been shown to possess such additive properties (306). There is no reason to believe, however, that operations or procedures will not be found with which to accomplish the objective. It may be, of course, that psychologists have been over zealous in their attempts to set up psychological scales which emulate physical scales, but it should be said that such attempts have been very

provocative in getting philosophers, physicists, and psychologists to analyze various problems of measurement.

The fact is that few psychological dimensions can be called extensive scales. In most cases psychological dimensions have simply ordinal properties and are called *intensive scales*. An ordinal or intensive scale is just a rank order scale in which A is reliably measured as being greater than B, B greater than C, and so on, but it is not known that, nor assumed that the difference between A and B is the same as the difference between B and C. If, for example, we were to rank three foods, T-bone steak, spinach, and cod liver oil in order of preference, we would probably rank them as written. However, our preference for T-bone steak over either spinach or cod liver oil is much greater than the preference for spinach over cod liver oil.

The Method

The method of equal appearing intervals has been used to break down a psychological dimension, or a portion of that dimension, into equal phenomenal units. In actual practice the *S* actively equates "sense distances." As an example, let us say we set up two lights, one bright and one dim; between these lights there is a certain brightness "distance." If you were given another light which could be varied in brightness, could you set the brightness of this variable light so that it would be halfway between the two standard lights? Could you bisect the brightness distance? If we gave you two numbers, say 5 and 11, and asked you to find the number which is half way between, you would easily establish this as the number *eight*. Could you, in the same fashion, find the brightness value which is halfway between the two standards—the psychological eight? If you could do this in a reliable fashion we would say that you had equated two sense distances. We might, then, give you two variable lights and ask you to trisect the distance between the standards, or three variable lights and ask you to quadrisect the distance. All of these procedures may be spoken of as *fractionation* of sense distances.

Another procedure which may be used has the qualities of multiplication. We present a standard light and a variable light and ask you to adjust the variable so that it appears to be twice as bright as the standard. We might also ask you to set the

variable so that it is half as bright as the standard. The latter becomes a form of fractionation since setting the variable half as bright as the standard would be bisecting the distance between the standard brightness and zero brightness.

From *S*'s standpoint, equal appearing intervals as a method describes itself very well. Equal appearing intervals are equal sense distances. In all the procedures, of course, many measurements are taken and measures of central tendency and variability determined.

Illustrative Experiments Using the Method of Equal Appearing Intervals

Working out an equal unit scale for a psychological dimension becomes a rather complex procedure. The problem is that of establishing equal psychological units throughout a dimension and relating each successive unit to a physical stimulus change. Since we know that the two dimensions are not perfectly correlated, we may expect that if psychological units are kept equal, the amount of physical change necessary to bring about these equal units must vary for different portions of the scale. We shall not attempt a complete follow-through on one of these projects, but we should go far enough so that you can get the idea of how such scales might be constructed.

A pitch scale. We shall use as our illustration the work of Stevens and Volkmann (367). The problem is that of constructing a pitch scale of equal units as a function of frequency, with all other aspects of the physical stimulus held constant.

Stevens and Volkmann used a special tone generator, an oscillator, which would present *S* a series of pure tones. Using the method of fractionation, *S* manipulated three variable tones to fractionate a given pitch distance into four phenomenally equal parts. One might think of this as bisecting first the total distance and then each of the two resulting halves, although in actual practice *S* was instructed not to work in this fashion. *S* was given as much time as he needed to make his equations, and several equations were made for each pair of standards.

We will consider the results obtained on one standard sense distance: namely, the distance between 200 and 6500 cycles. One standard was set at 200 cycles, the other at 6500. *S* was to

quadrisect this sense distance. A schematic drawing of the results on 10 *S*s is shown in Fig. 17. Each *S* made five judgments each. It can be seen that on the average the *S*s set the variable bisecting the total distance at 2022 cycles. The lower half was bisected at 867 cycles and the upper half at 3393 cycles. The four pitch distances resulting were equal from *S*'s standpoint.

Fig. 17. The quadrisection of a sense distance. *S*s were presented with two standard tones, 200 and 6500 cycles, and were asked to break the total pitch distance into four equal parts by setting three variable tones. *S*s judged 2022 cycles to bisect the total distance, and 867 and 3393 cycles the lower and upper halves respectively. Data from Stevens and Volkmann (367).

Now, how would we begin constructing a scale showing pitch as a function of frequency so that the units of the psychological scale are equal? First, we would plot the physical scale along the

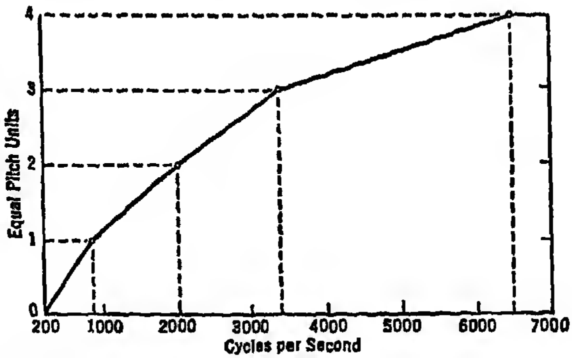


Fig. 18. Relationship between pitch and frequency in which the pitch units are equal. The zero of the ordinate is not to be taken as indicating zero pitch; it is a zero reference point for the sense distance used here. Data from Stevens and Volkmann (367).

abscissa. Usually this would be plotted in logarithmic units, but we shall plot it arithmetically as in Fig. 18. The pitch units, which the operations have defined as being equal, are plotted along the

ordinate so that equal ordinate distances represent equal pitch distances. Then if we coördinate the two axes as in Fig. 18, we have shown how much change in physical stimulation is necessary to bring about equal pitch changes along the dimension. Absolute zero can be determined by extrapolating the curve or by determining the lower absolute threshold. According to Stevens and Volkmann, both methods yield about the same result, showing the zero point of pitch to be at 20 cycles per second.

Results very comparable to those above have been found by presenting *S* with a single standard stimulus and requesting that a variable be set so that it will bisect the distance between the standard and zero. Using such a method, the following was found by Stevens and Volkmann.

Standard: (c.p.s.)	150	250	500	1000	2000	3000	5000	10,000
Frequency Judged Half:	85	111	206	373	633	1009	1437	2064

It is well to remember that a frequency judged half of a standard is a mean value, and successive judgments for the same standard would show a scattering of values. The fact that the variation is relatively small allows for rather precise scaling although, of course, the variability increases as the absolute frequency increases, as would be expected from Weber's law.

Other attempts to set up response scales of an extensive nature, as well as a thorough discussion of the problems attending those attempts, may be found in Reese's monograph (306).

Construction of an attitude scale. In all the experiments discussed thus far in Chapters II and III, *E* had control of a physical stimulus which was manipulated along a carefully quantified scale, with the major problem being that of relating a response dimension to the physical stimulus dimension. In the next chapter, and for the remainder of this chapter, we shall be concerned with problems in which *E* must attempt to quantify a *suspected* psychological continuum—a response dimension which is presumed to vary from high to low, from weak to strong, from pleasant to unpleasant, *but for which there is no known corresponding physical dimension.*

The method of equal appearing intervals, as well as other methods to be discussed in the next chapter, has been used in attempts to quantify human responses to stimuli which may vary

along many dimensions. We may ask, for example, to have *S* rate words for familiarity. Fairly high agreement could be reached on the familiarity dimension. What is the physical dimension which corresponds to the psychological dimension? Is it the number of letters in the word? Is it the angular shape of the letters? Is it the blackness of the ink? Is it a combination of these? Of course not. Familiarity can only seem to be very remotely related to such physical dimensions. To think of familiarity as apart from the judgmental responses made by *S* is like asking, "But, how many times *did* he stutter?" (cf. p. 47). We must be especially careful not to place the judgment in the stimulus—the judgments are *S*'s. No one has been more vigorous in cautioning against this than has Peters:

> It (the error) consists in explicitly or implicitly treating a stimulus variable which has no meaning apart from human reactions as if it had independent, physical existence. This is especially vicious and confusing in the field of aesthetics, where stimulus relationships and qualities, such as unity, complexity, simplicity, balance, symmetry, fusion, smoothness, formality, and representativeness, are studied in relation to value. Definition of qualities such as these is frequently given in terms of the physical object or its elements, or in a manner implying that anyone who took the trouble to look for them would remark their presence and amount.
>
> Stimulus qualities such as those listed . . . have no meaning except in terms of human reactions. They really *are* human reactions to the physical stimulus; and there is nothing within the physical stimulus itself which indicates the degree of the quality it will arouse in any one subject or in the average subject. . . .The most obvious measure of stimulus qualities is in terms of judgmental reports (30, pp. 274-275).

Our problem, then, concerns measurement techniques for responses to stimuli whose physical characteristics may seem utterly unrelated to the responses. The psychophysical methods may be adapted to accomplish this. Some of the most notable adaptations have been made by L. L. Thurstone. An illustration is his application of the method of equal appearing intervals to the construction of attitude scales. We shall work through one of his procedures in some detail, remembering that the purpose is to establish scale values for the stimuli where there is no physical dimension to be used as a reference continuum. For pitch we have cycles per second, for different colors we have millimicrons

of wave length but for attitudes there is no physical dimension which is correlated.

Thurstone and Chave (385) attempted to construct an attitude scale in which all units of the scale would be equal. The attitude chosen was the "attitude toward the church." It is assumed that this attitude extends along a conceptualized linear scale so that at one end we would find the most favorable attitude toward the church and at the other end the most unfavorable attitude toward the church. In the middle would be a neutral zone signifying indifference of attitude. The problem, then, is to quantify this supposed linear dimension so that its units are equal and so that a numerical index can be used in determining one's attitude toward the church once the scale is complete.

To do this, certain assumptions concerning attitudes and their expressions must be made. The definition of attitude given by Thurstone and Chave is typical of the usual definitions: ". . . the sum total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any given topic" (385, pp. 6-7). As indicated by this definition, and as most psychologists would agree, attitudes are extremely complex behavior tendencies. How are we going to measure them? The custom has been to use *opinions* as a means of measuring attitudes. The opinion is assumed to be the verbal concomitant of an attitude—it is a response which is assumed to be indicative of the nature of the man's complex reaction tendency toward a given topic. If a man says: "I believe in the church and all it stands for," we are likely to conclude that this man has a favorable attitude toward the church. The only other alternative we have is to believe that the man is a liar, which indeed will be true in some cases. This suggests that a man's behavior—his day to day activities—might be a better (or at least different) index of attitude than would his opinion. This may be true but getting a sufficient index of this behavior would be a laborious process. In lieu of this we may accept opinions as indices of attitudes and recognize that opinions and behavior are probably not perfectly correlated.

The first step in the construction of the scale was the gathering of a great number of statements which expressed opinions toward the church. These statements varied in terms of the opinions

expressed, from very unfavorable to very favorable toward the church. Thurstone and Chave collected 130 such statements. The second step was the actual scaling procedure by the method of equal appearing intervals.

Three hundred *Ss* or judges were given the 130 statements with instructions to sort them into 11 piles along the conceptualized linear scale of attitude toward the church. The judges were told that pile No. 1 indicated the end of the continuum which was most favorable toward the church and that pile No. 11 indicated the very unfavorable end, with pile No. 6 the neutral pile. The judges were instructed to sort each of the 130 statements into one of the 11 piles, depending upon the attitude which they thought was reflected by the written statement of the opinion. These 11 piles were to be thought of as being separated by equal sense distances. No judgment was made as to the equality of the distances; they were defined that way and so used in the statistical work.

The judges were not to express their own opinions toward the church; they were merely to sort the statements into the various piles, the sorting being done on the basis of their interpretation of the attitude implied by the statement. The judges were supposed to be impartial as far as agreeing or disagreeing with the statement. They were told that the piles need not have the same number of statements in them when the sorting was complete, but in most cases each judge used each pile at least once. Since 300 judges were used, each statement was allocated to a pile 300 times.

The raw data having been collected, considerable statistical work was required to refine the scale into a workable instrument. One of the most important considerations is that of measuring the degree of agreement among the judges as to the allocation of each statement. Suppose, for example, that a given statement were placed about equally often in each pile. This would indicate almost complete disagreement among the judges as to the opinion expressed by the statement; hence, as a means of tapping an attitude, it would be quite worthless. On the other hand, a statement which was placed in the same pile a great majority of the time would be indicative of high agreement among the judges. It becomes apparent that the variability of the placements is going to

be an important statistic for determining the *degree of ambiguity* of each statement. The greater the variability among the judges concerning the placement of a given statement, the more ambiguous we may conclude that item is.

The spread of the placements as indicative of ambiguity may be illustrated by the results obtained on two statements. The statement, "I don't believe church-going will do anyone any harm," proved to have a very low ambiguity rating as determined by the distribution of the placements. The median scale value of this statement was 5.8. On the other hand, the statement, "I believe the church has a good influence on the lower and uneducated classes, but has no value for the upper educated classes," shows almost a rectangular distribution, indicating little agreement among the judges. Figure 19 shows the distribution of pile allocations by the 300 judges for these two statements. The ambiguous statement, the one with little agreement, was placed in pile No. 5 by only 18 per cent of the judges and yet this was the greatest frequency in any pile. On the other hand, the first or unambiguous statement was placed in one pile (No. 6) by 62 per cent of the judges. There can be little doubt that this statement is less ambiguous than the other. Actually, there are several objective measures of the degree of agreement which might be used. The standard deviation would be satisfactory; the mean deviation might be used. Thurstone and Chave have used the statistic Q, which is the scale distance between the 25th and 75th percentile.

With scale values and Q values for all statements, the elimination of the undesirable items may begin. Ideally, of course, the final scale should have statements included so as to cover the entire attitude dimension, that is, so that the entire range of scale values is represented. For statements with very high and very low scale values we would expect, of course, markedly skewed distributions of judgments, but even in the case of those statements at the end of the dimension for which there is high agreement, the scale value would not fall at the very extreme. Our scale range must therefore be restricted somewhat as compared with the range with which the judges worked.

In making decisions as to which statements will be kept and which statements will not be kept for the final scale there is no

automatic rule to follow. There is no magic Q value above which statements are to be discarded and below which statements are to be retained. This is almost entirely up to E's judgment. Of course, the less the mean Q of the items in the final scale the more precise the measurement.

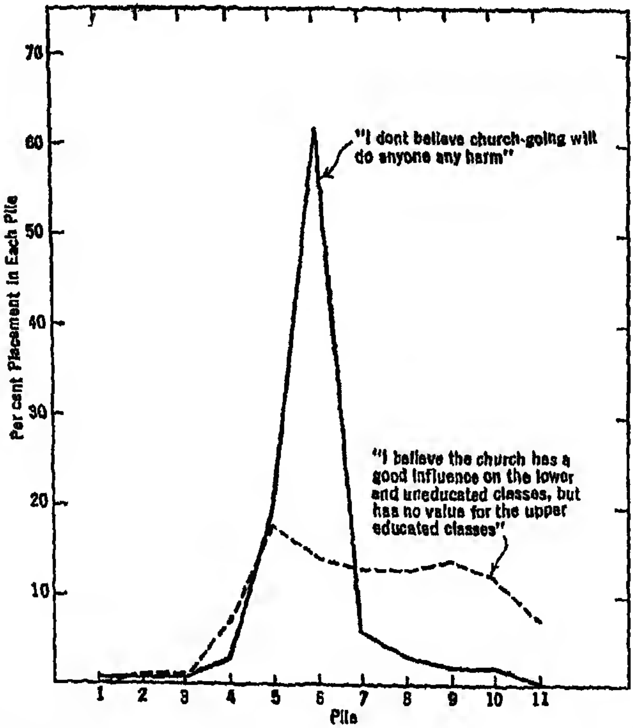


FIG. 19. An illustration of the difference between an ambiguous and unambiguous statement, in terms of the distributions of the pile placements. Data from Thurstone and Chave (385).

In determining which items to retain, we must also consider another factor. This factor is referred to as *relevance* of a statement. We might possibly get a statement on which there was high agreement among the judges in placing the item along the dimension but which was not very relevant to the attitude under consideration. An indication of the relevance of a statement may be obtained by requiring a group of Ss to check a tentative scale

by placing a plus before all the statements with which they agree. We can then determine a mean scale value for the endorsed statements—a measure of the attitude of a given *S* toward the church. Let us suppose that one *S* got a mean scale value of 7.0. Under such circumstances we would suspect that he would endorse nearly all statements which had a scale value of about 7.0, and that he would endorse with less and less frequency the statements with scale values further and further away from 7.0. Now, if the individual's mean scale value of the endorsements is 7.0, and we find an item with a scale value of about 7.0 which is not endorsed, we would be suspicious of that statement. The attitude it expresses may be irrelevant to the dimension. If enough *S*s showed such a trend for a given statement we would probably dispose of this statement as being irrelevant even though it might have a low ambiguity index.

In the final scale, after eliminating items because of high ambiguity and low relevance, Thurstone and Chave had 45 statements remaining, these statements being fairly equally distributed along the scale. It would be possible to make equivalent forms if enough "good" statements were found. We would also want to make checks on the reliability and validity of the scale. In the case of the scale constructed by Thurstone and Chave, reliability was high and indications of its validity were also obtained. For example, divinity students on the average gave scores which were further toward the "favorable" end of the scale than did regular college students.

It is well to point out that the method of equal appearing intervals is by no means the only method which can be used satisfactorily to construct attitude scales. There are others which appear to eventuate in scales as sound as those produced by the Thurstone technique and which at the same time require less labor (86, 220). Bird (26) gives a good outline of some of the various methods which have been used to construct scales. Our discussion of attitude scales arose because the construction of such a scale provides an illustration of the application of the method of equal appearing intervals.

SUMMARY OF THE METHOD OF EQUAL APPEARING INTERVALS

1. This method consists in "cutting up" sense distances into units which are psychologically equal.

2. With regard to sensory processes, the method has been used to establish precise relationships between the physical and psychological dimensions and to construct psychological scales which have the characteristics of physical scales. This latter has not been entirely successful.

3. Application of the method in other areas has been toward the equating of complex psychological distances for which there is no known paralleling physical continuum. The construction of an attitude scale was considered as an illustration of this.

4. Conceivably, the method can be applied to the establishment of equal psychological distances along any stimulus dimension.

CHAPTER IV

Discriminal Processes: Methods of Study III

In this, the final chapter on methods of studying the discriminal processes, we shall consider three methods: (1) the method of paired comparison; (2) the method of rank order; (3) the method of single stimuli.

THE METHOD OF PAIRED COMPARISON

The Method

This method is customarily used to determine relationships between *S*'s reponses for which there are no scaled physical dimensions. It may, however, be used in comparing stimuli which are physically measurable, and when so used it becomes quite similar to the method of constant stimuli. The procedure of paired comparisons is exactly what the name implies; *S* is presented two stimuli and is asked to judge which has the greater amount of a given characteristic. Each stimulus of a given group is compared individually with every other stimulus in that group. The stimuli may be simple, as would be the case when we determine pure color preferences, or they may be complex, as would be true when we ascertain preferences for works of art.

Let us presume that a manufacturer is going to market a new kind of pancake flour—something which he has never marketed before. He believes that part of the sales appeal of the flour will depend upon attractiveness of the package. Rather than trust expert judgment on the matter of package appeal, he has several different package designs made up and proceeds to determine empirically which of these is most preferred by a sample of housewives. If he has relatively few trial packages, say, 3 to 6, the method of paired comparison is ideally suited to get a solution to his problem. A sample number of housewives is asked

to compare each package independently with every other package as to attractiveness. From these comparisons he would be able to determine how much or by what proportion each package is preferred to every other package. Other things being equal, the manufacturer's final choice of package will be that one which is preferred over all others the greatest number of times.

We suggested that the method of paired comparison is a good one if there are not too many stimuli involved for comparison. Suppose that we had 50 packages for testing. The number of comparisons each person would have to make in order to compare each package with every other package is given by the formula, $\frac{n(n-1)}{2}$, n indicating the number of "stimuli" (packages). With 50 packages the number of comparisons becomes $\frac{50(49)}{2}$, or 1225 comparisons—a fair morning's work for any housewife. When a large number of stimuli is involved the method is not recommended because of the great amount of labor required in judging and in computation.

Let us consider another illustration. A sponsor of a radio program has three dance bands which he can get for the same fee. He considers it worth while to determine which of the three bands would be most preferred by the listening audience. Since there are only 3 "stimuli" involved, the method of paired comparisons may be used. A representative sample of the radio audience (the exact population might well vary as a function of the particular product which is being advertised) is presented with the names of the three dance bands, two at a time, with the question being somewhat as follows: "Which of these two bands would you prefer to hear in a 30-minute radio program of dance music?" With three stimuli, only three comparisons would have to be made, A with B, A with C, and C with B. From this hypothetical study the following proportions are derived:

A is preferred to B in 10 per cent of the cases
B is preferred to A in 90 per cent of the cases
A is preferred to C in 45 per cent of the cases
C is preferred to A in 55 per cent of the cases
B is preferred to C in 85 per cent of the cases
C is preferred to B in 15 per cent of the cases

From these data we would have little difficulty in concluding that dance band B is likely to be the most attractive to the greatest proportion of the population. If that proportion represents potential buyers, of the product to be advertised, the sponsor will probably be wise to choose band B over the two rivals considered. It is clear, of course, that such data do not indicate that B is the most attractive of all bands—it only indicates that of the three compared, B would be most preferable to the greatest proportion of the population sampled.

We may ask the question: "How much more do people prefer B to A or C?" We have a relatively large "distance" between either A or C and B, and only a short distance between A and C. The "how much" can be answered, although the complexity of the procedure does not recommend its inclusion here. The problem is essentially the one we have discussed in the previous chapter in establishing equal unit scales. This can be done if certain assumptions are granted. The mathematically inclined student can find the assumptions and the problems worked through in several sources, though the most recent and lucid expositions are given by Thurstone (384) and by Gulliksen (126). For our purposes it is sufficient to know that rank order of preference may be obtained from paired comparisons, and that in addition there are techniques for establishing the psychological distance separating the various ranks.

Illustrative Experiments Using the Method of Paired Comparisons

Judgment time in aesthetic choices. A study by Dashiell (75) illustrates the work on determining preferences for pure color and at the same time demonstrates the relationship between time required to choose between the two colors and the psychological distance which separates those two colors. We have already touched upon this problem in the case of the judgment of length of lines under the conditions of the method of constant stimuli. Dashiell's problem was: "Does the time taken to make a choice vary consistently with the amount of difference in the effective values of the alternative stimulus-objects?" (75, p. 57).

Seven colors were chosen from a standardized series. The colors were red, green, yellow, blue, orange, orange-yellow, and

purple. They were presented to *S* by a tachistoscope and introduced in the following manner:

> You are to look through the two openings in the box before you. You will see first a gray card with two dots. Fixate these. When I press my key, the dots will go away and two colors will appear. I want you to tell me which of these colors you prefer. Here are two keys. Place the index finger of each hand on a key. You are to indicate which color you prefer by the key that you press. Press the key which is on the side of the color you prefer. When you make your choice, press the key firmly. It is not necessary to hit it. You must make a choice, but there is no hurry. This is *not* a reaction-time experiment (75, p. 59).

Actually, however, *E* did measure reaction time. This was accomplished by *E*'s pressing a key when the colors were shown, thereby starting a clock which was stopped when *S* pressed a key indicating a choice. Two cycles of the experiment were run so that the space error would be cancelled out. (You will remember that the space error was considered in Chapter II in discussing the Müller-Lyer illusion experiment, and was defined as a bias which *S* might have for objects to the left as compared with those to the right, or vice-versa. With two cycles in the present experiment, the colors were presented equally often on both sides.)

Dashiell's results showed that in this sample of college students blue was most preferred and orange-yellow least preferred. The complete rank order as derived from the proportions of preference for each color over every other color was, from most preferred to least preferred: blue, green, purple, red, orange, yellow, and orange-yellow. The latency of response was determined by calculating the mean reaction time between each given color and the colors of all other ranks. For example, when blue was judged with the various other colors the mean latency was 1.39 seconds. When red was judged with the other colors the mean latency was 1.84 seconds, a longer mean reaction time than in the case of blue. The least-liked color, orange-yellow, was responded to negatively, but more rapidly than those in the middle of the series. These findings are similar to other findings which we considered under the method of constant stimuli and seem to be typical of the judgment process. When differences

are clear-cut there is little hesitancy by *S*; when the differences are slight, hesitation occurs, thus increasing the judgment time. This principle probably holds for all human behavior, whether it concerns the judgment of differences in colors, weights, political issues, new clothes, or occupations.

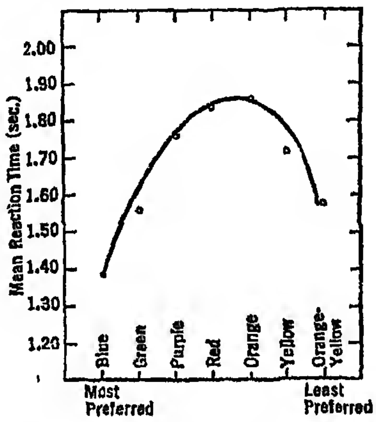


FIG. 20. Relationship between color preference and reaction time in choosing between two colors. The mean reaction times were determined by calculating the mean reaction time in choosing between a given color and all other colors. Nine *Ss* were used in determining the mean values. Data from Dashiell (75).

It will be noted in Fig. 20 that preferred colors tend to give somewhat more rapid reaction times than do the least preferred colors. This finding has been substantiated by a carefully scaled paired comparison experiment of Shipley and his students (353). Pursuing the problem further, Shipley (354) performed an experiment in which *S* was instructed to choose the color liked *least* in the successive pairs. Under these conditions the results were reversed—*Ss* showed shorter latencies in choosing between the least preferred colors and the other colors. Shipley interprets this finding as being indicative of a factor of set brought on by the instructions and not as intrinsic to affective judgment as such. *S*, instructed to choose the *preferred* color, had a *polarity of set* toward preferred colors. This set enhanced speed of reaction. *S*, instructed to choose the *least preferred* of a pair of colors, had a set toward reacting more rapidly toward the least

preferred end of the series. As we shall stress again and again, the importance of the instructions to *S* cannot be over-emphasized. Minor variations in instructions—perhaps supplied unwittingly by *E*—may, in certain instances, strongly influence the results. Unless the instructions constitute the independent variable, they should be scrutinized as rigorously as all other factors in the experimental situation.

Paired comparisons in the study of values. Barker (9) has used a modified form of the method of paired comparison on a rather unique subject matter. College students were presented with various terms or phrases descriptive of traits and values which might be desired in one's living. Here are some of the phrases: *good health; cheerful disposition; good humor; gaiety; strong religious faith; financial security.* In all, there were 18 such phrases, and since each phrase was paired with every other phrase, there would be a total of 153 pairings or judgments. These phrases were presented on mimeographed sheets and *Ss* took the sheets to their homes, with instructions not to make all the judgments in one sitting. Barker did not ask *S* to choose between two simple alternatives, but rather, between paired alternatives. For example, rather than ask *S* whether he would prefer *good health* or *strong religious faith,* Barker requested him to choose between the two alternatives, (1) good health, with less than average religious faith, and (2) strong religious faith, with less than average health.

Such a procedure presents a rather complex task to *S*. Four positions (two on each of the two dimensions) must be weighed simultaneously. That is, *S* must know or conceive of the meaning of good health and below average health and at the same time balance these against his conception of strong religious faith and below average religious faith. The complexity of the task was probably further heightened by certain statements which might suggest more than a single dimensional characteristic. For example, one of the alternatives offered for choice was *artistic ability or appreciation,* which to some *Ss* may suggest two quite different values. In short, *S* was making choices between exceedingly complex alternatives.

In addition to the above procedure, Barker had *Ss* indicate whenever their choices were uncertain. The instructions read:

In some cases your choice will be obvious and clear to you. In other cases your choice will be very uncertain, and you will have to force yourself to indicate a choice; in these cases draw a circle around the *x* (9, p. 43). (*S* used *x* to indicate his choice.)

From the results of these comparisons Barker rank-ordered the 18 values, from the one receiving the greatest number of choices to the one receiving the least. (The actual rank orders of the values are not given.) He found that the greater the difference between the attractiveness of the two alternatives the less the likelihood that the choice was marked as uncertain. He also found that if two alternatives were both very attractive, the choice was more likely to be marked uncertain than if the two alternatives were both quite unattractive.

In a previous study (8) Barker had shown that when the attractiveness of two stimuli is about equal, *S* shows greater vacillation, or, as it is sometimes called in describing both human and animal behavior, vicarious (substitute) trial-and-error behavior. All data, therefore, seem to point toward several basic behavioral concomitants of difficult choices: (1) vacillatory behavior, (2) long reaction times in making a choice, and (3) higher uncertainty of the "correctness" of the choice.

Preferences of broadcast listeners. A study by Eisenberg and Chinn (87) was aimed at determining what tonal range and intensity is most preferred on radio broadcasts. The experiment was performed in a specially fitted room at the Columbia Broadcasting Company studios in New York. The room was furnished to make it much like the living room in an ordinary home. *S*s, in most cases, were obtained by advertising for them over the key CBS station in New York. This method of obtaining *S*s, while not insuring a representative sample, does at least represent a landmark in the history of psychology. Although *E*s have used a great variety of techniques to obtain *S*s, this was probably the first time that radio was used for this purpose.

"Tonal range" in this study refers to degree of completeness of tonal range, as one aspect of the fidelity with which an original program of music or speech is reproduced when broadcast and picked up by a particular receiver. Radio receivers do not, as a rule, reproduce faithfully all the tones which occur, say, at a symphony concert. Three tonal ranges were used in this

experiment: wide, medium, and narrow, as determined by the number of high and low frequencies which were cut out in the transmission. Thus, the wide range of tonal reproduction extended from about 30 cycles to 10,000 cycles per second, while the narrow range extended from about 100 to 6,000 cycles per second. The wider the range the greater the number of overtones or harmonics which are transmitted. In special conditions, three different intensity levels were used, these being 50, 60, and 70 decibels.

The transmitted program consisted of music and speech, in most cases recorded, though in one instance a regular program was fed into the testing room. With three different variations in tonal range, three judgments are necessary to complete the pairing of each stimulus with every other stimulus. Also, with variations in intensity three additional pairings are necessary, and if the two variables are worked out in relation to each other, several more would be required. This latter was not done completely.

Each comparison involved 1 minute. During this minute two conditions were presented alternately for 10 seconds each, so that *S* actually heard each three times during the 1 minute period and before he had to state a preference. Obviously, *S* could not have both stimuli presented simultaneously and make a valid judgment, so the alternation scheme was used. *S* was notified of a change at the end of each 10-second interval by a signal light. As in the other experiments we have discussed, *S* was asked to state a preference for one or the other stimuli, but could, if necessary, indicate "no choice."

The results of this study showed that in listening to music, *Ss* actually preferred a narrow band of tonal qualities rather than the wider bands. This is somewhat contrary to expectations, in that it has usually been assumed that the radios which will reproduce most faithfully are the ones which will be preferred. It is customary for salesmen to extoll the virtues of their radios by emphasizing that the speaker will pick up very low and very high notes, and that this makes for better listening. The present results suggest that a wide range of reception is not necessarily preferred. That the results are not a function of the *Ss* obtained by advertising is suggested by the fact that a group of trained

musicians also had a distinct preference for the narrow tonal range of musical stimuli. In the case of transmitted speech, the medium tonal range was preferred over either of the extremes. Intensity levels most preferred averaged between 60 and 70 decibels, with the somewhat higher level being preferred for speech.

Summary of Method of Paired Comparison

1. The method of paired comparison is a method for studying the discriminal processes by comparing every stimulus in a given set with every other stimulus with regard to a given dimension. The method is most efficiently used when the number of stimuli is small, since many stimuli make the procedure laborious.

2. The method has been applied to a great variety of stimulus characteristics, both simple and complex. It is possible to derive rank orders of the stimuli or, by making certain assumptions, to determine scale values which show not only the ranks but also the "distance" between successive stimuli.

3. While the method is applicable to stimuli which can be meaningfully described by a physical dimension, most of the current applications of the method involve stimuli for which there is no scaled or otherwise kind of physical dimension.

4. The method is closely related to the method of constant stimuli and to the method of rank order—the next method to be considered.

THE METHOD OF RANK ORDER

The Method

This is one of the simplest methods used to study the discriminal processes and correlated behavior, and in many respects it is the most efficient. *S* is presented a group of stimuli which are presumed to have some dimension in common and is requested to rank the stimuli in order from high to low according to the characteristic being considered. The rank assigned a given stimulus by all *Ss* may be combined by calculating a mean rank. As in the other methods, measures of variability should be obtained, and, as in the case of paired comparison, scale values may be determined by certain statistical manipulations. The method is usually applied to the study of judgments or discriminations in

which there is no known physical stimulus correlate to the response evoked.

Due to the similarity of the problems for which the method of paired comparison and the rank order method are suited, two illustrations should suffice.

Illustrative Experiments Using the Method of Rank Order

Nationality preference. Eighty-five members of an elementary psychology class were given slips of paper on which were listed seven different nationalities: Japanese, Chinese, Russian, French, Italian, English, and German. The class was instructed as follows:

> Imagine that you move into an International House on a college campus. One of the rules of the House is that you must have as a roommate a foreign student of one of the seven nationalities listed. Look at the list carefully, then assign the number *one* to the nationality of the student you would most prefer to have as a roommate; the number *two* to the nationality preferred second, and so on, until the least preferred nationality is assigned the number *seven*. Take as long as you like to make the rankings. If two nationalities appear to be a tie as far as you are concerned, choose one or the other—don't rank them the same. All seven nationalities must be assigned a rank and no two ranks are to be the same.

In preparing the slips, many different orders of the nationalities had been made so that no bias should attach to the serial listing of a given nationality when all *S*s were considered. The mean ranks assigned to each nationality were determined from the data. Distributions of the rankings for nationalities rated near the bottom or top were skewed as one would expect. The mean ranks and the standard deviations of the distributions are shown in Table 6.

TABLE 6

NATIONALITY PREFERENCE DERIVED BY METHOD OF RANK ORDER

Nationality	*Mean Rank*	*Standard Deviation*
English	2.23	1.36
German	2.75	1.66
French	3.22	1.63
Russian	4.19	1.85
Chinese	4.42	1.66
Italian	5.14	1.33
Japanese	6.04	1.36

It is immediately apparent that the method of rank order gives about the same results as those obtained by the method of paired comparison. In one sense, both methods can be considered the same. When *S* ranks, he should, ostensibly, compare each stimulus with every other stimulus if his ranks are to be assigned carefully. If he did this he would be following the method of paired comparison. The virtue of the paired-comparison technique over the method of rank order lies, if at all, in the fact that by paired comparisons *S* is *forced* to make independent judgments of each stimulus with every other stimulus. *S* may not actually make such a thorough analysis in the case of rank order. Nevertheless, it has been shown empirically that the two methods will yield results which are essentially similar (149). Rank order is preferable if the number of stimuli to be judged is large.

Olfactory preferences. This study will demonstrate a modification of the method of rank order and will also touch on a sensory field which thus far we have said little about. *Ss*, in this study by Eysenck (92), were asked to rank in order olfactory (smell) stimuli. The stimuli were presented to *S* in small corked bottles with only a letter on each bottle as identification symbol for *E*. The olfactory sense adapts very quickly, so *E* must not present successive stimuli too rapidly. In this study 31 different stimuli were used, and rather than ask *S* for a complete ranking of so many stimuli which would be hard to make, *E* first instructed *S* to place the stimuli in three groups according to the judgment: *definite pleasantness*, *average pleasantness*, and *definite unpleasantness*. After putting all stimuli in one of the three groups, *S* was then asked to refine his judgments by giving each stimulus a score according to the following arrangement:

Score:	10	9	8	7	6	5	4	3	2	1	0
No. of stimuli to be given the score:	1	2	2	3	4	7	4	3	2	2	1

This procedure forces *S* to put the stimuli into a symmetrical distribution. The best-liked (most pleasant) odor was given a score of 10; the next two in order of preference were given a score of 9, and so on, until the most unpleasant odor—the rankest—was to be assigned zero. The method thus combines rank order with a rating scale made up of forced frequencies in each interval.

We need not list all the stimuli—many of them would be familiar only to chemists. Aromatic stimuli, such as vanilla, cinnamon, and eau de cologne were usually ranked as the most preferred odors, while asafoetida and pyridinal were indicated as being the most unpleasant.

Summary of Method of Rank Order

1. With this method *S* is asked to place in rank order a group of stimuli according to the amount of a given characteristic. Mean ranks may be used as the basic statistic showing the relationships among the responses evoked by the stimuli.

2. This method is closely related to the method of paired comparison and is usually used in lieu of that method when many stimuli are involved.

THE METHOD OF SINGLE STIMULI

The Method

This method was mentioned briefly when considering the method of constant stimuli in the preceding chapter. We are discussing it more fully at this point because most of the experimental work with this method has employed stimuli for which there is no explicit physical dimension to relate to the responses. By the method of single stimuli *S* is presented a single stimulus at a time and is asked to make a judgment concerning it. He may be asked if the stimulus is "pleasant or unpleasant," "acceptable or unacceptable," "good or bad."

The relationship of this method to the method of constant stimuli is seen in the fact that when *S* makes a judgment by the method of single stimuli, his standard is "within himself," whereas in the case of the constant method used to measure difference thresholds the standard is presented along with the variable. The present method is sometimes called the *method of absolute judgment*, implying that there is no other stimulus to which *S* can refer his judgment. In the sense that there is no other stimulus manipulated by *E*, this is true, but as we shall see, subjective standards may be built up quickly and may retain considerable permanence.

The method of single stimuli has certain advantages which

none of the methods discussed thus far has. Let us go back for a moment to the illustration given when discussing paired comparisons. Here we presented *S* with three stimuli—names of dance bands—in pairs and asked for a preference at each pairing. Now, unless we pursued this problem further we might, in certain instances, find that none of the three bands was highly pleasing to the listeners. If the sample of listeners had been presented with the bands (figuratively) one at a time and asked: "Do you think a thirty-minute concert by this band would be *enjoyable* or *not so enjoyable*," it could conceivably be found that none of the three bands would make good listening. Similarly, if we are going to study the aesthetic responses toward certain modern works of art, a paired comparison or a rank order need not necessarily give us the answer to the question of whether or not the work actually elicited approval or disapproval responses.

Large business concerns, being sensitive to changes in public attitude toward them, may have regular public opinion surveys taken so that a close record of shifts of opinion (if any) may be obtained. Changes can be noticed following certain events, such as strikes. These surveys, designed to keep the company abreast of public reaction, often use the method of single stimuli. The person being interviewed will be asked: "Do you think *well*, or, *not so well* of United Steel?" Do you think *well*, or, *not so well*, of General Brands Corporation?"

The method, then, consists of presenting *S* with successive stimuli and requesting a judgment of these stimuli along a specified dimension. The data can be handled readily. For any given set of stimuli we can determine the proportion of *Ss* using specified descriptive categories for each of the stimuli.

Illustrative Studies Using the Method of Single Stimuli

Affective "Conflict." Emotional reactions have been favorite responses for study by the method of single stimuli. The human organism makes affective responses to a wide variety of stimuli. The responses may be only vague, undifferentiated feelings, or they may be clear-cut and violent responses, such as rage or fear. Many of the so-called affective responses resolve themselves into no more than what are commonly called likes and dislikes. The responses, if described in strict behavioral terms, are approach

responses for "likes" and avoidance responses for "dislikes." In most cases, of course, the overt movement indicating approach or avoidance is short-circuited and a verbal expression of "like" or "dislike" is all that is measurable. Even here, however, there may be observable physiological changes which suggest emotional responses. Not only do our responses become telescoped but also the stimuli which are capable of eliciting emotional responses may, under certain circumstances, be transferred to the word which stands for them. There is, for example, nothing intrinsically unpleasant about the word *vomit*, but most people will rate such a word as being quite unpleasant.

The study of affective responses of different intensities to verbal stimuli is a fertile field of experimentation for psychologists. We will consider an exemplary study of the affective responses to a group of common words as determined by the method of single stimuli. This study is reported by Lanier (209).

Ss were given the following instructions to read before the experiment started:

> This is an experiment in "affective judgment." The purpose is to study the feeling tone produced in different individuals by the same common words. The experimenter will call out a word, and you are to report at once how the word affects you, using one of the following terms: PLEASANT UNPLEASANT INDIFFERENT MIXED
>
> Use the category MIXED when a word seems to arouse both unpleasant and pleasant states. Make your judgments immediately. Do not try to analyze or to rationalize your reactions. Simply report the immediate affective reaction to the words (209, p. 204).

Ss were 38 students at Vassar College. In addition to obtaining the number of times each word was placed in the four categories by the method of single stimuli, Lanier measured reaction time to the stimulus and the psychogalvanic response. The psychogalvanic response is a much studied reaction in a great variety of laboratory investigations, although there is no general agreement on its specific interpretation. It is agreed that this response is indicative of emotional arousal and is brought about through the activity of the autonomic nervous system. The response is measured by recording changes in electrical resistance of the skin with a device called a galvanometer. *Ohms*, a unit of electrical resistance, is the common response unit.

We should comment on the MIXED category used by *E*. It is possible that we do have mixed reactions toward certain stimuli. We may both like and dislike the event or object symbolized by a word, or at least we may have approach tendencies toward certain portions of the stimulus complex and avoidance tendencies toward other portions. Clinicians have recognized these opposed tendencies and the term *ambivalence* has been used to describe the condition. A person, it is alleged, may both love and hate his father, or to be more accurate, we would say that some of the father's behavior evokes a withdrawal response and some evokes an approach response. The MIXED reaction should, therefore, be clearly differentiated from the INDIFFERENT response, the latter being a neutral state with neither pleasant nor unpleasant reactions —the stimuli evoke no affective responses.

TABLE 7

REACTIONS OF 38 *S*s TO A GROUP OF COMMON WORDS

	Pleasant	*Unpleasant*	*Mixed*	*Indifferent*
Percentage in Each Category	50.2	24.4	10.0	15.4
Median Reaction Time	1.79	1.89	2.73	2.38
Median Psychogalvanic Response	165	177	263	166

Data from Lanier (209).

Table 7 shows the percentage of times each of the four categories was used to indicate the response to the verbal stimulus. Also shown are the median reaction times and the median psychogalvanic response for the words. It is to be noted that only 10 per cent of the words elicited a mixed affective reaction. However, for these reactions the latency of response was longer and the psychogalvanic response was greater than for any of the other categories. The differences between the reaction times and psychogalvanic responses of the mixed category and those of the other categories are significant far beyond chance expectation. If the psychogalvanic response is indicative of an emotional reaction, the data suggest that the affective change brought about by the mixed words was greater than it was for the words in which the affective responses were clear-cut (either pleasant or

unpleasant). On the other hand, it is difficult to see why the indifferent group of words elicited such large psychogalvanic responses. Nevertheless, to Lanier the long reaction times and the large psychogalvanic responses indicate that a miniature "conflict" was taking place during the judgmental process concerning the mixed words. Table 8 shows the 10 words most frequently

TABLE 8

TEN WORDS MOST FREQUENTLY JUDGED PLEASANT, UNPLEASANT, INDIFFERENT, AND MIXED

Mixed	*Indifferent*	*Pleasant*	*Unpleasant*
body	brick	father	worry
work	street	love	war
study	finger	mother	pain
drink	habit	song	blood
money	cousin	home	fear
pray	tone	marry	cry
figure	cut	dance	dead
hot	tree	peace	sick
book	grass	swim	sin
kiss	body	water	crush

From Lanier (209).

placed in each of the four categories. Note that there is only one repetition—the word *body* occurs in the first two lists. *S*s thus showed high agreement in the placement of the words. There can be little doubt that words can be reliably sorted according to feeling tones evoked by them.

An anchoring effect. In any situation in which *S* is requested to make judgments concerning a given characteristic of a series of stimuli, the range of the characteristic will probably influence the judgments. It is as if *S* sets up a descriptive dimension of his own and judges the various stimuli according to that subjective dimension. This dimension is determined not only by the range of responses to the stimuli being employed, but also by the repertoire of responses which *S* brings to the situation. For example, *S* may judge certain of the stimuli as being unpleasant, but that does not necessarily mean that the stimuli being used represent his conception of the most extremely unpleasant stimuli possible. Unpleasantness and pleasantness are not to be

thought of as non-dimensional in nature; there are degrees of pleasantness and degrees of unpleasantness among one's reactions. It can be shown that a given subjective dimension which *S* uses to evaluate a given series of stimuli can be shifted as a consequence of the instructions which are given him. An experiment by Hunt and Volkmann (173) will demonstrate this.

The method employed by Hunt and Volkmann makes use of the method of single stimuli but adds a feature which makes the problem comparable to that used in the well-known rating scale procedure. A rating scale employs the method of single stimuli but defines the characteristics of the dimension on which the judgments are to be made more thoroughly than is customary in the method of single stimuli. Rating scales are used widely to evaluate performance of employees or trainees in industrial situations. They were also used to evaluate performance in military service during the late war. In the usual rating scale procedure an instructor or supervisor is asked to rate *S* along a scale such as follows:

Excellent	Very Good	Good	Average	Poor	Very Poor	Failing

Numbers might be assigned arbitrarily to the various points; descriptive terms might be limited to the extremes of the dimension, and several scales might be used to rate various aspects of the individual's performance. There are many possible variations on the rating method. Like any measuring device, the rating scale should meet the requirements of consistency or reliability.

Hunt and Volkmann did not present their *Ss* with a rating scale as such, but instructed them in the first part of the experiment as follows:

> You will be shown a series of colors. You are to judge the pleasantness and unpleasantness of each color on a scale of seven steps, represented by the numbers *one* to *seven*. Let the higher numbers stand for the pleasant end of the scale, and the lower ones for the unpleasant end. Report your judgment to the experimenter as soon as it is made (173, p. 89).

Thus, *Ss* were not told that the most pleasant color in the series was to be assigned the number *seven*, nor the most unpleasant

color *one*, but neither were they told not to do this. Five *Ss* followed these instructions, rating each of 10 colors 10 times. Ten trials gave *Ss* plenty of opportunity to get a good conception of the entire range of stimuli being employed. When they completed this series of judgments, *Ss* were instructed as follows for the second series:

> You will be given a series of colors. You are to judge the pleasantness and unpleasantness of each color on a scale of seven steps, represented by the numbers *one* to *seven*. Think of the most pleasant color you can; then hold the pleasantness of this color in mind throughout the experiment, and let its pleasantness define the step *seven* on your scale. When you report "seven," you are to mean a pleasantness approximately equal to the pleasantness of the color which you are holding in mind. Report your judgment to the experimenter as soon as it is made (173, p. 89).

Note that these instructions suggest to *S* that there may be more pleasant colors than any represented in the 10 colors shown. *Ss* are instructed to think of the most pleasant color they know, and hold that in mind for evaluating the colors actually presented to them. This "thought-of color" thus becomes the anchoring point around which the other judgments will hinge to a certain extent. As Hunt and Volkmann predicted, under the second instructions there was a shift downward of the judgments in the upper part of the scale. This shift is shown in Fig. 21. *Ss*, under the second instructions, no longer gave the most preferred color a value of nearly seven; rather, the average rating of the most preferred color dropped to about 5.6. This "stretching" of the subjective scale occurs only in the upper regions, i.e., among the highly pleasant colors. Had *Ss* been instructed to think of number *one* as being the most unpleasant color they could conceive, it would be predicted that the dimension of pleasantness and unpleasantness would be further stretched out, because the actual ratings of the 10 colors would show a still further restriction in range of the placements.

The influence of the anchoring value as produced by instructions to *S* is that of enlarging the frame of reference within which the judgments are made. *Ss*, unless specifically instructed, will set up their own frame of reference. Here (in the second part of the Hunt-Volkmann study) a frame of reference is imposed on

one end of the dimension by providing a specific anchoring point around which the other judgments are to be ordered.

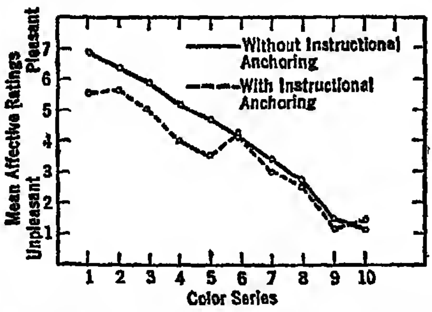


FIG. 21. An anchoring effect in the judgment of the affective tone of colors produced by specific instructions. The affective scale is along the ordinate, with a rating of one indicating the unpleasant end and seven the pleasant end. The data are mean values on 5 *Ss*. Data from Hunt and Volkmann (173).

It was predicted above that if Hunt and Volkmann had used an anchoring instruction concerning the low (unpleasant) end of the scale, the data would have suggested a further expansion of the dimension. This prediction has been confirmed in another experiment by McGarvey (230) using a completely different kind of material. In a portion of this experiment, *S* was asked to rate a series of "behaviors" along a six-point scale from most desirable to least desirable. These behaviors were such things as: *fishing without a license, poisoning a neighbor's dog whose barking bothered you, killing an idiot baby, manufacturing counterfeit money,* etc. Twelve such behaviors were used in the main experiment and *S* rated these items along the six-point scale. On a second series of ratings, *Ss* were provided with anchoring points at *both ends* of the scale, the anchoring behavior in one case being less desirable, and in the other, more desirable than any of the 12 behaviors which had previously been rated. The influence of this procedure was to cause a considerable restriction in the range employed to rate the 12 behaviors the second time. It was as if the anchoring behavior produced a lengthening of the dimension so that, relatively, the 12 rated behaviors seemed less extreme than formerly.

Summary of the Method of Single Stimuli

1. In this method *S* is presented with successive stimuli and is asked to judge or rate them on a given attribute. The attribute might be a fairly complex one (such as degree of pleasantness) or a simple one (amount of "heaviness"). The method may be used with stimuli for which there is or is not an explicit physical dimension, although the latter is usually the case.

2. The method employs the basic rationale of all rating scale devices.

3. *S* may construct his own subjective standard around which his judgments are made. Because of this, the method has many similarities to the method of constant stimuli in which a real standard is used. In several studies the subjective standard has been changed by instructions which introduce a stimulus (real or symbolic) to represent a standard value at the end of the dimension.

SOME CONCLUDING CONSIDERATIONS

In these three chapters (II, III, and IV) we have been concerned with stimulus-response relationships which were found in *S* upon his coming to the laboratory. This does not mean that his responses are not modifiable, for in the next chapter we shall see that under certain conditions even simple sensory relationships may be changed. What does seem apparent, however, is that in the experiments which we have considered, *E* was not concerned with bringing about relatively permanent changes in the discriminal processes by manipulating environmental conditions. The methods of studying the discriminal processes were developed in a scientific environment interested largely in measuring the processes as they existed at the moment and not how they came to be as they are at that moment. Only recently has there been a definite shift in interest toward the study of the *development* of our discriminal capacities. Only lately also has there been a trend toward the application of the methods to the study of a far greater range of problems than those for which they were devised. It now appears that the methods are unusually well suited to the study of social, moral, political, and economic problems.

We must state again that in reviewing the methods of experimentation in the past three chapters we have only scratched the surface in *giving the factual results* which have emerged from the application of the methods. Although the sound factual data on sensory and perceptual relationships are enormous in amount, it is not the proper and specific purpose of this book to review them. If in another book they were presented merely as facts they might appear to be a hodge-podge of unintegrated knowledge, but if they were systematized around *theory* they would take on significance. Such a systematization would require at least a volume each for audition and vision. Such books would include considerable information concerning the structure of the sense organs and the nerve tracts associated with them, since, by and large, theories of sensory functioning are physiological in nature. The facts upon and out of which the various theories have been erected have been obtained largely by the use of the methods outlined in the three chapters, especially Chapters II and III. The references at the end of this section may be consulted for these facts.

Another point should be reëmphasized. Without the many researches which have been performed in developing the methods of study that have been outlined in these three chapters, much of the research on the more distinctly psychological level (such as the measurement of attitudes) might not have been possible. The men who for hours on end have lifted weights and listened to barely perceptible sounds have provided us with a variety of techniques which are useful in a great many situations. It is a fair guess that long after the facts derived from psychophysical research have been assimilated into the general body of knowledge, historians will say that the major contribution made by psychologists in the realm of the discriminal processes was the development of research techniques which were applicable far beyond the immediate problems on which they were worked out.

We should, by now, be clearly aware that even the most elementary form of experimentation requires careful and persistent control. By now, too, we should have a basic grasp of the problems involved in experimental design, so that in succeeding chapters we may become, where necessary, more critical of some of the work and claims of experimental psychologists. It will be

made very clear that ability to outline and carry through an experiment requires a profound appreciation and respect for a multitude of factors or conditions which might bias or falsify the results. And our final conclusion will be that control of the multiplicity of conditions which may make for constant errors in even simple experiments—conditions which to the uninitiated might seem inconsequential—demands a vigilance not normally forced on us in other pursuits.

SPECIAL REFERENCES

The following references may be consulted for systematic factual and theoretical information concerning the stimulus dimensions which influence our discriminal processes.

1. *Psychology for the Fighting Man*, by a committee of the National Research Council (Washington: *The Infantry Journal*, 1943).

This volume provides one of the most readable accounts available of sensory and perceptual relationships. Special emphasis is given to the practical application of these facts to soldiering. Chapters II through VII are most pertinent.

2. Boring, E. G., Langfeld, H. S., and Weld, H. P., *Foundations of Psychology* (New York: Wiley, 1948). Chapters 10 through 17.

The treatment of sensory and perceptual relations in this book is more detailed than other texts written for the elementary student. Each chapter is written by a specialist in the given area. A good introduction to theories of sensation for various sensory modalities is given.

3. Woodworth, R. S. *Experimental Psychology* (New York: Holt, 1938).

Although intended for advanced students, the treatment in this book is so lucid that a beginning experimentalist can handle much of the material. Chapters XVII through XXVI.

CHAPTER V

Discriminal Processes: The Influence of General Variables

INTRODUCTION

We have mentioned previously that the discriminal capacity of the organism is dependent upon (1) physiological structures which are sensitive to certain energy changes, and (2) practice in making discriminations. The importance of the first factor is fairly evident; if the retina of the eye is destroyed we are insensitive to light waves; if the spinal tracts carrying the nerve impulses from the muscles of the legs are destroyed, as sometimes happens in general paresis (severe syphilitic infection), there are no kinesthetic cues to aid in the "direction" of the leg. In general, soundness of the physiological mechanism is a prerequisite for making fine discriminations. The second factor, the influence of practice, is distinctly more psychological. When we say that the discriminal processes are influenced by practice in making discriminations we are saying that *learning* serves to influence our responses to the physical energies in the environment.

The importance of learning in the development of our discriminal processes is usually granted in general but seldom specified in detail. Most researches in the past have been concerned with cross-sectional analyses of the discriminal processes, i.e., the capacity of the organism at the moment. When we study the influence of learning on these processes we are interested in the longitudinal aspects, i.e., how they developed. We have pointed out that practice effects are usually present even in a short experimental session, and that in cross-sectional studies, *E* must control these effects so that his data are not biased as a result of a practice differential. Early in this chapter we shall look at

experiments in which *E*, rather than controlling the effects of practice (as by counterbalancing), studies these effects. The bulk of this work has been done during recent years and the results have demonstrated clearly that our conception of the responses which are subject to modification must be broadened.

Another factor which influences the discriminal processes, and which is basic to learning, is the factor of *motivation*. Our interest in motivation as related to discriminal processes has been sharpened because a few researches show that the motivation of *S* at the time he is making a discrimination or a judgment will profoundly influence that discrimination.

One word before we review these experiments. A more thorough coverage of learning and motivation takes place in later chapters. There we will systematically detail the experimental manipulations which cause changes in motivation, and also those manipulations which bring about differences in the rate of learning. By so doing we shall give adequate operational definitions to the concepts. For the present chapter we shall leave the two concepts of motivation and learning undefined, believing that their generally understood meanings are sufficient to permit an interpretation of their influence on the discriminal processes. When certain researches apparently show that factors other than learning and motivation influence the discriminal processes, the best guess is that these hitherto unidentified factors are actually aspects of learning and motivation. However, because of the unanalytical nature of the studies we shall have to be content with references to as yet unanalyzed biases influencing the discriminal processes.

Our three major subject matter sections for this chapter, then, will be the examination of a few illustrative studies which show the influence on the discriminal processes of (1) learning, (2) motivation, and (3) unanalyzed biases. In addition to the review of these subjects, attention must be paid to certain problems of experimental design. Thus far in the book we have been dealing largely with experiments in which the estimate of the influence of a given condition was determined by the mean of many measurements on only one *S*, or at most, a few *S*s. In some of the experiments to be reported in this chapter and in most of the experiments in the remainder of the book, we shall be dealing

with procedures involving a *single measurement on many Ss.* The problems of experimental design raised by these procedures are numerous. Our initial consideration of these problems occurs in the latter part of the present chapter, using for illustrative material experiments dealing with the discriminal processes.

ILLUSTRATIVE STUDIES SHOWING THE INFLUENCE OF LEARNING ON THE DISCRIMINAL PROCESSES

Pitch discrimination. We know that the major physical stimulus correlate of pitch is frequency of the sound wave. We know also that there are easily recognized differences among individuals with regard to pitch discrimination. The person who sings off key with irritating abandon probably has poor pitch discrimination. The problem of the influence of training on pitch discrimination has been a subject of considerable controversy for many years, with some psychologists asserting that pitch discrimination was largely a "given," not subject to change, while others thought of it as a capacity which in part at least was built up by learning and would change with adequate training. Studies undertaken to resolve this difference of view had been quite inconclusive until Wyatt (423) performed an experiment in which most of the shortcomings of the previous studies were eliminated.

Wyatt used the Seashore Measures of Musical Talent as the major index of pitch discrimination. Two other measures were also used, but we shall not consider them here. The Seashore Measures, it will be remembered, get an index of musical talent by employing the method of constant stimuli, in which the stimuli are presented in pairs on phonograph records. The album includes 12 sides, among them a test of pitch discrimination. A standard tone is presented with variable tones 1, 2, 3, 5, and 8 cycles above and below the standard. An average *S* will have only a small percentage of error when the difference between the standard and variable is 8 cycles, and will have nearly a chance score (50 per cent correct) when the difference is 1 cycle. The less the difference between the standard and the variable which can be perceived, the better the discrimination.

Wyatt used 16 *Ss*, 8 of whom are described as musically inclined (being enrolled in the School of Music) and 8 as not

being musically inclined. The 16 *Ss* were administered the Seashore Measures for Pitch twice. This procedure provided a stable index of initial proficiency in pitch discrimination. Then each individual *S* was given twelve 50-minute periods of training in pitch vocalization and pitch discrimination. After the training they were tested again on the Seashore Measures. The pre-training tests were given at the start of a semester and the post-training tests at the end of that semester. No control group was used. It is a likely assumption that at least in the case of the non-musical *Ss*, little change would be brought about in pitch discrimination by other activities which took place during the semester. Such an assumption is not completely justified in the case of the 8 *Ss* enrolled in the School of Music.

The training procedures were somewhat flexible, being changed to suit the needs of each *S*. Most of the discrimination training took place with a standard tone at 500 cycles, though in a few instances a tone somewhat lower than this was used. A special training device, a *stroboscope*, was employed. A tone of 500 cycles was sounded and *S* was instructed to reproduce the pitch vocally. If the vocal tone were off pitch (from the 500-cycle tone), this would be registered automatically by the stroboscope. Thus, *S* could tell immediately whether his voice was too high or too low.

TABLE 9

INFLUENCE OF PRACTICE ON PITCH DISCRIMINATION

	Difference Between Standard and Variable (c.p.s.)				
	8	5	3	2	1
Pre-training error (%)	16.7	28.6	31.3	38.3	49.8
Post-training error (%)	0.3	9.4	25.4	26.9	42.1

Data from Wyatt (423).

Improvement in pitch discrimination was shown in the improvement between the scores on the initial and final test by the Seashore Measures. These data for all *Ss* are shown in Table 9 and Fig. 22. The average increase in the discriminal ability is pronounced at all levels of difficulty, i.e., with both fine and gross differences between the standard and the variable, but the

greatest percentage of improvement took place with the coarser discriminations. The non-musical *S*s scored higher *after* training than the musical *S*s did *before* training. After training, 13 of the 16 *S*s were above the median for the group on which the Seashore Tests were standardized; before training, only 5 *S*s. The changes brought about in these *S*s by Wyatt's training program leaves no doubt that pitch discrimination can be significantly improved by proper training. It is not known from this study whether or not the improvement which took place was permanent. If it follows the pattern of other learned acts, we may expect forgetting to take place unless practice is continued.

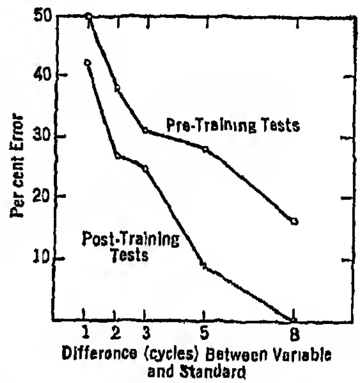


Fig. 22. The influence of training on pitch discrimination. Data from Wyatt (423).

Wyatt's study (and others, e.g., 312, 401) leads us to conclude that pitch discrimination is not a fixed capacity with which we are endowed, but is amenable to change by learning. The increase in the ability to discriminate differences in pitch as a function of training undoubtedly has its counterparts in other psychological dimensions of auditory phenomena, as well as in other sense modalities. In the broader conception of our discriminal processes, Moore's (264) classic study has shown that what we hear as consonance and what we hear as dissonance in musical tones is probably a function of the learning imposed by our cultural standards. There is nothing inherently dissonant or consonant about a given combination of tones; rather, certain combinations are called dissonant, and they are "bad," while others are said to be harmonious or consonant, and they are "good."

Affective judgments. A series of experimental reports by Peters has been most effective in demonstrating the rôle of learning in our affective responses to certain stimuli. Peters starts out with an ideal situation by using stimuli to which *S*s had no initial strong

affective response. By laboratory manipulation he then attempts to produce different feelings toward the stimuli. In one study (288), Peters used 49 Japanese words. These words would not produce strong affective reactions in *S* because for the most part they were completely new stimuli. Some affective tone might be aroused, however, because of similarity to English words. Sixty *Ss*, using the method of single stimuli, rated all 49 words as being either pleasant or unpleasant. From the original 49 words, 10 were chosen for special treatment. They were put on a *memory drum* (a device for exposing verbal material for short, regular intervals). *S* was instructed to pronounce 5 of the words whenever any one of them came into the drum window and *not* to pronounce the other 5. Thus, 5 of the stimuli were to be followed by an "approach" or positive response, while 5 were to be followed by an "avoidance" or negative response. *Ss* continued working on the 10 words until they had gone through the list five consecutive times without error.

TABLE 10

INFLUENCE OF FORCING POSITIVE OR APPROACH RESPONSES AND NEGATIVE OR AVOIDANCE RESPONSES TO VERBAL STIMULI

	Per Cent Times Judged Pleasant	
Positive Words	Before Training	After Training
warui	38.3	70.0
mushi	43.3	71.7
tori	43.3	73.3
yoi	56.7	81.7
heiwa	56.7	91.7
Negative Words		
kenasu	60.0	51.7
karai	53.3	35.0
tatakai	50.0	30.0
homeru	70.0	48.3
amai	78.3	61.7

Data from Peters (288).

After the learning criterion had been achieved, *Ss* were asked to judge the stimuli again as to pleasantness and unpleasantness. Table 10 shows the initial and final scores in terms of the

percentage of cases a given word was rated pleasant. The first 5 words in the table were those *Ss* had to learn to pronouce, and the second 5 are the ones they had to learn not to pronounce. In every case where *S* had to learn to make a positive response—to pronounce the words—the pleasantness value went up, as shown by the number of *Ss* rating them as such. In every case where *S* was forced to avoid pronouncing a word, the pleasantness index went down. These results are squarely in line with Peters' theoretical argument that affectivity is a learned response and that stimuli which evoke avoidance responses will be judged unpleasant and stimuli which evoke approach responses will be judged pleasant. The meaning of the stimulus, then, determines the affective judgment toward that stimulus. It would thus appear that if we can produce a stimulus which evokes withdrawal responses, the feeling of unpleasantness will develop toward that stimulus; contrariwise, if we can produce an approach response toward a stimulus, the feeling of pleasantness will develop toward that stimulus.

Attitude changes. Thurstone, in applying psychophysical methods to the study of social problems, has shown in several instances that changes in attitudes can be brought about by setting up certain experiences which are designed either to counter or enhance a given attitude. One illustration will suffice (384). This study used the paired comparison technique and school children as *Ss*. Initially, the children scaled the seriousness of various behaviors. *Ss* were presented the stimuli in pairs and were asked to indicate which of the two they thought most serious. There were 13 items involved, with the children placing a gangster, a bank robber, and a kidnaper near the top in terms of seriousness of offense, and putting the beggar and the tramp near the bottom or least serious end of the scale. A gambler was placed near the middle of the scale. After the initial comparisons, the children were shown a film which depicted the life of a gambler in a none too favorable light. Then the children again evaluated the 13 offenses by the paired comparison method. The only appreciable change occurred in the case of the gambler, and the seriousness of this offense was raised considerably. Evidently the film was effective in producing a change in attitude toward gambling. This study exhibits the influence of learning

since the acquisition of the "true" picture of the gambler, as portrayed by the film, raised gambling to a higher point on the dimension.

Experimental establishment of a perceptual bias. Extensive data presented in Chapter II, Table 2, showed a consistent tendency for *S* to overestimate the length of a given line. Such a tendency is indicative of a perceptual bias. The Müller-Lyer illusion provides another illustration of a perceptual bias which is almost universal. How do such biases arise? There is no agreement on this matter. The universality of certain illusions and their presence at early ages suggest that learning is not an important determinant of them. Learning may, however, modify these perceptions considerably. On the other hand, certain of our other perceptual biases seem to have been developed almost exclusively by learning. It is to such biases that we will turn our attention.

One of the difficulties attending the investigation of such problems is that many of the biases are so unique that general trends are difficult to discern, and even if general trends are isolated, not a great deal is told us about how the bias developed. The source of the bias may be far back in the life of the individual. Yet, there is no reason why these perceptual biases could not be developed in the laboratory if they are learned. If they are learned, the laboratory problem consists in applying basic principles of learning to discover what causes *S*'s perceptual bias. We turn to a study by Proshansky and Murphy (300) which suggests that this can be accomplished.

Two small groups of *S*s were used, a control and an experimental group. In a pre-training series of judgments by the method of single stimuli, *S*s of both groups were asked to estimate the length of a series of lines from 2 inches to 7 inches long. Then the experimental group was put through a series of training trials. During training *S* was shown the same series of lines he had seen during the pre-training trials. In the training trials, *S* made no overt response to these stimuli—he only observed them. However, *E* rewarded and punished *S* according to a prearranged plan. All long lines were rewarded by giving *S* money when they were shown to him; all short lines were punished by *E* taking back some of the money. (At the end *E* was nine dollars in the hole.)

Intermediate length lines were rewarded and punished in a haphazard manner. The control group went through the training series but were merely shown the lines—no rewards or punishments were given.

The critical point of the experiment came in the post-training test series. On the test series *S*s of both groups were shown intermediate length lines and were asked to estimate the absolute lengths. No rewards or punishments were given to either group during the test series. The statistics involved the comparison of the pre-training judgment of the lines with the post training judgments. This comparison showed that there was a significant shift in the experimental group toward estimating the length of the lines *longer* than they had been estimated during the pre-training trials. The control group showed no such shift. From these results it is inferred that the experimental group had developed a perceptual bias toward overestimating the length of lines as a consequence of being rewarded for the longer lines. The duration of the bias was not determined.

Modification of taste preferences. We all have our preferences for certain foods. Some of the food likes are almost universal; an American who does not like ice cream is looked upon with astonishment. Many other foods are liked by some and disliked by others in the same fashion that certain musical selections or works of art are liked by some and not liked by others. Manufacturers of new food products make it almost routine to determine the best formula (as determined by a taste test in a sample of the population) for a given product. If one is to can a new soup and there are several combinations of ingredients which might be used, taste tests may be used to determine which combination would be most preferred. The armed services, as well as most large food manufacturing concerns, maintain testing laboratories for the purpose of testing tastes. Their problem is to determine preferences as they exist. An illustrative study will show how these preferences are likely to be the product of learning.

Gauger (104) set up a procedure which is called *conditioning.* The purpose of the investigation was to determine whether the food preference of pre-school children could be changed. He concocted some distasteful liquids made up of vinegar, egg whites,

and salt. Children, when forced to ingest these solutions, showed obvious displeasure. Judges rated the reactions of the children on a dimension labelled "satisfaction" at one end, and at the other, "dissatisfaction." In the initial part of the experiment the solutions of vinegar, egg whites, and salt were rated as bringing about very great dissatisfaction on the part of the *S*s. Then a training procedure was put into effect whereby a piece of chocolate, a highly satisfying food, was given immediately after the distasteful food. These pairings were continued for 35 days, with the judges rating the children's reactions to the foods each day. By the end of 35 days, the previously rated dissatisfactory foods were rated as no worse than indifferent, i.e., they brought neither satisfaction nor dissatisfaction. The chocolate, on the other hand, fell markedly in the amount of satisfaction expressed, though by the end of the 35 days it had not reached the point of indifference. Each food seemed to have influenced the degree of liking or disliking for the other as a consequence of the situation which forced the ingestion of both liked and disliked foods in rapid succession. The distasteful food became less distasteful and the tasteful food became less tasteful. These conclusions depend, of course, on the assumption that the reactions of the children as rated by the judges are a valid index of food preferences.

Summary. In this section we have considered experimental illustrations which all point to the importance of learning as a determinant of discriminal processes. It has been shown that even in simple sensory discriminations, such as pitch difference thresholds, as well as in our discriminal processes which result in what are called attitudes and affective judgments, learning is a major determinant of the response. Learning thus becomes a fundamental variable in discriminal processes, and an understanding of factors which effect the rate of learning, as well as allow it to take place, becomes basic to the understanding of behavior.

ILLUSTRATIVE STUDIES SHOWING THE INFLUENCE OF MOTIVATION ON THE DISCRIMINAL PROCESSES

The psychological size of coins. One of the most intriguing experiments to be published in recent years is one reported by Bruner and Goodman (45) in which 10-year-old *S*s estimated

the size of coins. The method used was the method of average error. Twenty *Ss* served in an experimental group and 10 in a control group. The apparatus was a small box with an adjustable photo diaphragm mounted at one end. The diaphragm could be adjusted so that a circular patch of light, admitted through the opening, could be varied in size from $^1/_2$ to 2 inches.

The experimental group of *Ss* went through two series of judgments. In the first series *S* was asked to set the diaphragm so that the circle of light would be the same size as a penny. Then he was asked to make the opening the same size as a five-cent piece; then a dime, a quarter, and finally, a half-dollar. For each coin in this series *S* made two judgments, one after *E* had set the diaphragm considerably larger than the coin, and one after *E* had set it considerably smaller than the coin. When the size of all coins had been estimated once, the procedure was repeated, but this time *S* started with the half-dollar. The exact procedure for each *S* was:

1¢	*5¢*	*10¢*	*25¢*	*50¢*	*50¢*	*25¢*	*10¢*	*5¢*	*1¢*
S-L	*S-L*	*S-L*	*S-L*	*S-L*	*S-L*	*S-L*	*S-L*	*S-L*	*S-L*

The *S* and *L* indicate "smaller" and "larger," showing that for each coin in both orders the diaphragm was set both smaller and larger so that the movement error, if present, would not unduly bias the judgments. The ascending and descending order (*a* to *b*, *b* to *a*) would presumably balance any practice effects which might occur. Each *S* went through the above procedure so that each made a total of four judgments for each denomination of coin.

In the first part of the experiment *Ss* set the diaphragm on the basis of memory for the size of coins. After the completion of this series they went through the same procedure again except that a real coin was held in the palm of the left hand while adjustments of the diaphragm were made with the right. Thus, *S* would hold successively a penny, a nickel, and so on, in his hand, in each case adjusting the diaphragm until he thought it was the size of the coin.

A control group went through only the second series, and instead of holding a coin in their hands, the *Ss* held a small cardboard disc. A series of these discs was cut so as to be exactly the same size as the series of coins held by the experimental group.

If there were differences appearing between the judgments of the two groups it is presumed that they were a function of differences in *apparent size* and not *actual size*.

So far we have discussed conditions for gathering the following data: (1) judgments when coins are absent; (2) judgments when coins are held in hand; and (3) judgments when cardboard discs are held in hand. The selection of the 20 *S*s in the experimental group was such that half came from well-to-do families and half from a settlement house in a slum area. Differences in judgments between the poor and rich children thus permit study of the effect of another variable.

TABLE 11

PERCENTAGE DEVIATION FROM ACTUAL SIZE IN JUDGMENTS OF COINS AND DISCS

Group and Conditions	*Penny*	*Nickel*	*Dime*	*Quarter*	*Half-Dollar*	*Number Judgments Per Coin*
20 *S*s Coin Present	16.5	23.9	29.1	37.0	29.6	80
20 *S*s Coin Absent	7.2	19.6	11.6	32.8	35.8	80
10 *S*s Disc Present	—5.4	—.9	—1.5	1.8	—.8	40
10 Rich *S*s Coin Present	10.3	20.4	16.3	22.4	17.4	40
10 Rich *S*s Coin Absent	2.6	19.8	7.8	28.3	34.7	40
10 Poor *S*s Coin Present	22.7	27.3	41.8	51.6	42.0	40
10 Poor *S*s Coin Absent	11.8	19.4	15.4	37.3	36.9	40

Data from Bruner and Goodman (45).

Table 11 shows the results of the judgments for all conditions for all groups. The data are presented in terms of percentage deviation from actual size of the coins or discs. All values which have a minus sign indicate the *S*s set the diaphragm smaller than the true size of the coin, and all others indicate that they set it larger than the true size. From Table 11, one general result is

quickly perceived; whenever the judgments were made concerning the coin size, overestimation of the size resulted.

Figure 23 is a graphical representation of the judgments made by the 20 *Ss* in the experimental group when the coin was present as compared with the judgments made by the 10 *Ss* in the control group when the cardboard discs were used. In judging the discs the control group was remarkably accurate except for the 5 per cent underestimation of the disc the size of a penny.

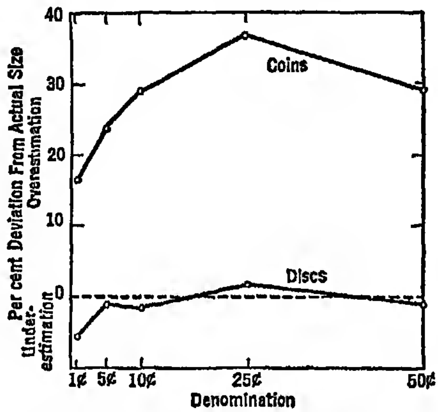


FIG. 23. Errors in the judgment of size of coins and size of cardboard discs the same dimensions as the coins. Data from Bruner and Goodman (45).

The experimental group, on the other hand, consistently set the diaphragm so that it was larger than the true size of the coin. This overestimation is large and except for the drop which occurred with the half-dollar, the error increases as the *value*—not size—of the coin increases. In the case of the quarter the overestimation averaged 37 per cent. A quarter has a diameter of about 15/16 inches. A 37 per cent error indicates that on the average the diaphragm was set so that its diameter was about 21/16 inch.

Table 11 indicates that except in the case of the half-dollar, the overestimation was not as great when the judgments were made from memory (when coin was absent) as when the coins were present. Also, this table shows that the overestimation of the coins by the 10 *Ss* in the poor group is greater than that made by the 10 *Ss* in the rich group. These latter differences have been plotted in Fig. 24.

Bruner and Goodman indicate that these differences in the magnitude of the overestimation between the poor and rich children probably reflect a basic motivational difference in the two groups. A coin has greater value to a poor boy than it does to a rich boy and this is presumably mirrored in the size estimations.

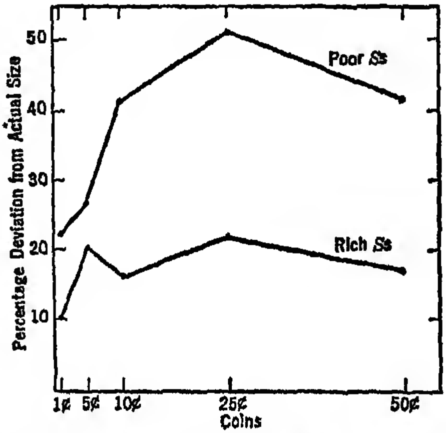


FIG. 24. Differences in the amount of overestimation of coin size as a function of the financial status of *S*. The poor *Ss* were from a settlement house in a slum area, the rich *Ss* from well-to-do families. Data from Bruner and Goodman (45).

The conclusion of the authors with respect to the over-all results of their experiment is that in perceptual judgments one must consider motivational factors which may operate to distort the "pure" perception. If a given object has reward value—if it would satisfy a motivational state—observations concerning that object may be distorted, and a constant error results. From the results of this experiment it appears that the amount of constant error is a direct function of at least two variables: (1) the strength of the motivation (rich *Ss* vs. poor *Ss*), and (2) the potential reward value of the object (the greater the value of the coin the greater is the constant error, within limits).

Concerning the procedure used in conducting the experiment, several factors should be pointed out. A cardboard disc would not weigh as much as a coin the same size; hence, this may be a variable not controlled. The coins have designs on them, the

cardboard discs do not. This might influence the size estimations. The rich and poor groups were not matched for perceptual acuity, the assumption being that the average perceptual acuity of the two groups was the same. Since the obtained differences were large, it is unlikely that the control of these factors would change the basic findings, but the introduction of such controls might change the absolute differences somewhat.

Response modification by hunger. Although a fear of hunger may serve to motivate many people, hunger as such is probably not a vital factor in most of our adult behavior. This would not be true, of course, among inmates of concentration camps nor among very poor people. However, it can be shown that lack of food for even short intervals of time will modify the response characteristics of adults. Two experiments which demonstrate this have been performed by Sanford (331, 332). The experiments are closely integrated and we shall take the essentials from both for our considerations.

The question which Sanford asked was: "How do food responses vary in frequency as time since food was taken increases?" To answer this question, *S*s were brought to the laboratory at various times following a meal and were subjected to certain tests designed to elicit food responses. One of these tests was a *free association test.* Selected words were used as stimuli to which *S* was instructed to give the first response to the stimulus word which occurred to him, i.e., the first response which came to his mind. Some of the stimulus words used were: *finish*, *plenty*, *store*, *hot*, *knife*, *serve*, and *bottle*. All of these words, it can be seen, might be associated in some manner with eating. If the response to a stimulus word was the name of a food or a verb meaning *to eat*, it was counted as a food response.

Another method of getting an index of the change in frequency of food response utilized what is called commonly a *projective technique. S*s were given pictures and asked to tell what the pictures suggested to them. If there were variations in the number of responses clearly indicating concern over food, and if these variations were related to the time interval following the last meal, influence of the basic motivational state of hunger would probably be indicated.

Although other measures of the frequency of food responses

were used, the two discussed above were the major measurements taken. The results of the experiment show that food responses did increase as the length of time between testing and the last meal increased. The greatest increase in the number of these food responses took place during the first few hours, with but little change between 4 and 24 hours. The mean number of food responses for various intervals is shown in Fig. 25. The mean values are composite scores of all methods of measuring the food responses.

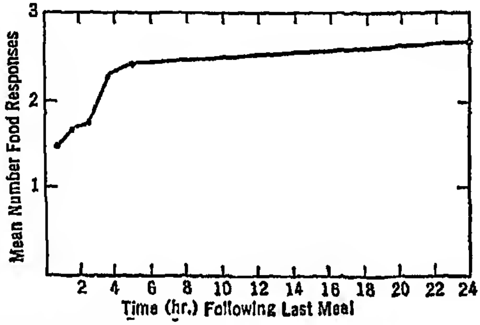


FIG. 25. Relationship between the number of food responses and time elapsing since last meal. Data from Sanford (332).

A somewhat similar experiment has been performed by Levine, Chein, and Murphy (214). *E*s showed 5 *S*s a series of achromatic (non-colored) cards and chromatic (colored) cards on which were ambiguous or indistinct figures as seen through a ground-glass screen. *S* was requested to report what he saw in the screen. Like Sanford's results, it was shown that up to a certain point the number of reports of seeing food increased as the time since last eating increased. However, beyond 6 hours for the achromatic pictures, and beyond 3 hours for the chromatic pictures, the number of food interpretations decreased. It is doubtful if this decrease is reliable. McClelland and Atkinson (228) have recently employed very much the same technique as that used by Levine, Chein, and Murphy, but used from 17 to 20 *S*s in various groups. The longest period of deprivation for *S*s was 16 hours, but up to that point no evidence of a decrease in the number of food responses was found.

Perceptual differences brought about by different motivations induced by hypnosis. Next to extra-sensory perception, probably no other psychological phenomenon has caused so much furor as has hypnosis. Whereas extra-sensory perception is of dubious authenticity, hypnosis is universally recognized as a genuine phenomenon, differing from the normal waking state. Descriptively, hypnosis is a state of heightened suggestibility, although the basic mechanisms which allow an individual to be placed in an hypnotic state are unknown. Hypnotic states are of all degrees of intensity, from extremely light hypnosis differing but little from normal consciousness, to deep hypnosis in which *S* is literally a tool for *E*. Because of the sensationalism attached to the phenomenon, because of misunderstandings concerning it, and because it has been employed by unscrupulous individuals, it is still looked upon with disfavor on many college campuses. The major use of hypnosis at present is in the psychotherapeutic treatment and diagnosis of maladjusted individuals. If care is taken, however, there is no reason why hypnosis may not be used as a laboratory technique. We shall evaluate one experiment in hypnosis as an exercise which will strengthen your critical attitude of experimental procedure. See if you can detect the flaw in the experiment to follow before it is pointed out.

The technique used consisted of inducing different motivational states in deeply hypnotized *S*s, and determining the differences in responses which resulted in a standard situation. In this experiment only 3 *S*s were used. Each *S*, while hypnotized, was "given" in succession first a happy motivational state, then a critical motivational state, and finally, an anxious motivational state. While each *S* was under each of these three different states, a series of six pictures was shown him and he was asked to describe or interpret what he saw in each picture. *E*s were primarily concerned with the different interpretations which emerged as a function of the different motivational states. By proper suggestion under hypnosis the memory of the previous description for a given picture under one state can be removed so that when *S* describes the same picture under a different motivational state, it appears to be an entirely new picture. You may find it interesting to ponder this fact with regard to the balancing of practice effects.

It will be of value to indicate the instructions used in inducing the three motivational states:

HAPPY. Now you are feeling very happy and you are in a cheerful and joyous mood. You feel as if everything is rosy and you are very optimistic. You have a comfortable feeling of well-being; nothing is worrying you. You feel perfectly at peace with everything and everyone. You are in a very happy, cheerful, and optimistic mood.

CRITICAL. Now you are very critical; you are quick to find fault and to condemn unfavorably. Your judgment of others is very harsh and severe. You see failings and faults very clearly. You are very critical and fault-finding.

ANXIOUS. Now you are quite anxious. You are disturbed over some possible misfortunes. You are disquieted and concerned as to something in the future. You are a little fearful and mildly alarmed. You have a feeling as if you were expecting something disagreeable to happen, yet were not sure it would. You are quite anxious (213, p. 518).

E recorded everything *S* said when describing the pictures. The core of these remarks has been reproduced in the report. The remarkable differences in the interpretation given the pictures under the different motivational states can only be demonstrated by reproducing some of the remarks. Let us take one picture and quote the remarks made by all 3 *S*s under three different motivational states. This picture depicted 4 college men on a sunny lawn. These men were typing, reading, and listening to the radio. The following interpretations were given:

State	*Subject A*	*Subject B*	*Subject C*
HAPPY	"Before the war. Boys lying on campus listening to radio and reading. Not getting much done but having a good time. That's fun though."	"Complete relaxation. Not much to do—just sit, listen, and relax. Not much at all to think about."	"Sunbathers, radio, books, studying and typing in the sunshine. It must be a very warm and restful place."
CRITICAL	"Typical college campus. Everybody loafing. Typing a letter he doesn't need to type. Slacking on their work."	"Someone ruining a good pair of pressed pants lying down like that. They're unsuccessfully trying to study."	"Somebody trying to get tan, wasting time. Not comfortable. Tie coming undone, sole of shoe worn out. Messy. Leaves unraked."

State	*Subject A*	*Subject B*	*Subject C*
ANXIOUS	"Boys relaxing on college campus. By now they may be overseas or wounded or dead. They never know whether they'll come through alive."	"They're listening to a football game or world series. Probably a tight game. One guy looks as if his side wasn't winning."	"Lot of men. Peaceful enough. I wonder if they want to come back, and what's going on in their minds. Are they fighting now?" (213, pp. 519-521).

Further, an attempt was made to see how much the induced states influenced the thoughts or perceptions of the *S*s as inferred from the remarks. Each remark made by *S*s for every one of the six pictures was typed on a separate slip of paper and given to three judges who were requested to decide whether the remarks indicated a happy, critical, or anxious mood. In the case of the remarks produced under the happy states and the critical states, the judges were in high agreement; that is, remarks produced during these states were placed in the proper category by the judges. Concerning the anxious state, however, the results were much less clear-cut. The judges sometimes thought the remarks obtained under this induced state were critical and sometimes thought them happy. This is apparent even in the protocols reproduced above.

The results obtained tend to support the importance of motivational differences as a factor producing differences in perception. Indeed, as the authors point out, in several instances the description given one picture in one state would not be recognized as applying to the same picture for which a description was given in another state. It seems that *S*s see different things, although the environment remains the same.

Can we accept the conclusion indicated? Doesn't the experiment possibly contain a false inference with regard to the influence of hypnosis in producing the differences? It would appear that a worthwhile control to this experiment would have been to have *S*s *not* hypnotized describe the pictures following the same instructions given to the hypnotized *S*s. It is barely possible that non-hypnotized *S*s would have described these pictures in as diverse ways as did the hypnotized *S*s. To accept the results as indicated we have to assume that normal *S*s would have behaved differently. Such assumptions should not have to be made.

Summary. We have considered experiments which *suggest* the importance of specific motives in determining the direction of the discriminal processes. This is a relatively new area of research and the data are far from being conclusive for discriminal processes as a whole. Enough has been done, however, to make two points highly probable: (1) motivation influences what for some time has been thought to be relatively invariable response patterns; and (2) the dimensions which describe a physical stimulus are far from being the only variables which enter into a psychological situation in which *S* is asked to make judgments concerning that stimulus. Physical dimensions may be used to describe the stimulus as such, and variations in those dimensions will usually bring about a variation in response. However, in addition, we have the accumulated past experience of *S* with the stimulus, plus the motivation at a given moment, which together may in some cases alter the discriminations drastically.

ILLUSTRATIVE STUDIES SHOWING THE INFLUENCE OF UNANALYZED BIASES ON THE DISCRIMINAL PROCESSES

In reviewing most of the studies in this section we are likely to speculate concerning the origin of the biases which will be shown to influence the discriminal processes. Usually the proposed hypothesis will point to learning as the basic determinant. Yet, the complexity of these biases, and the lack of control of the learning in the laboratory, make the attempt to trace the source rather like a guessing game.

Biased judgments as a consequence of wearing glasses. Many people hold a fairly firm belief that they can judge quite accurately the personality traits of an individual merely by looking at the person or perhaps talking with him for a few minutes. The number of judgments which is made in this manner every day is legion. Only after studying the results of experiments such as the one by Thornton (382) can we fully appreciate how much these everyday judgments may be based on trivia which have little if any pertinence in making a reliable or valid judgment.

Thornton asked: "What differences in judgment of the personality traits of a person might arise as a consequence of his wearing or not wearing glasses?" To answer the question, Thornton

set up two conditions: In the first, *S*s wore glasses and were judged on several traits by a group of students; in the second, the *same* *S*s did not wear glasses and were judged for the same traits by students. By this procedure each *S* became his own control so that the results of the judgments when he wore glasses could be compared directly with the results of the judgments when he did not wear them. Such a procedure requires, however, that the same person not appear before the judges both with and without glasses, since the judges would undoubtedly remember the individual as having appeared before. To get around this difficulty Thornton divided the two classes of elementary psychology students (who served as judges) into two sub-groups. *S* then appeared before one sub-group wearing glasses and before another not wearing them. The same procedure was repeated for the second class. Eleven *S*s were brought before the two sub-groups of the first class and 10 of these *S*s appeared before the two sub-groups of the second class. Schematically the conditions appear as follows:

	CLASS A			CLASS B	
	Sub-group I	*Sub-group II*		*Sub-group III*	*Sub-group IV*
5 *S*s	With	Without	5 *S*s	With	Without
6 *S*s	Without	With	5 *S*s	Without	With

The *with* indicates that *S* wore glasses; *without* means *S* appeared without glasses.

In making his appearance before the groups of judges (37 in Class *A* and 29 in Class *B*), *S* came into the classroom and sat in a chair for approximately 2 minutes. The glasses which were worn under the appropriate conditions were rimless and made of plain clear glass. The *S*s were graduate students and were unknown to the group of judges. The judges were told nothing of the purpose of the experiment other than it was designed to determine how accurately people can judge traits by visual cues. The judges made estimates of several personality traits including *kindliness*, *intelligence*, *industriousness*, *honesty*, *dependability*, and *sense of humor*. The judges recorded their estimates on a 10-point rating scale on which zero meant "completely lacking in the trait"; 5 meant having an "average amount," and 10, having the "greatest possible degree." The rating scales were treated arithmetically and

significance of the differences in the ratings was found in terms of the differences in the mean ratings obtained.

The results for all ratings combined show that *when wearing glasses S was judged to be more intelligent than when not wearing glasses.* The difference between the two means was highly significant. Furthermore, while wearing glasses *S* was likely to be judged as more *industrious*. The difference in effect between wearing and not wearing them was not as great as in the case of intelligence, but large enough to be reliable. Judgments of the other traits showed no significant relationship to the wearing of glasses although several approached significance. However, all traits were judged to be present in somewhat greater amounts when *S* was wearing than when not wearing glasses except *sense of humor*. Here the judges tended to bestow a higher sense of humor on *S* when not wearing glasses. In another study (381) Thornton has shown that all the differences in the judgments were greatly enhanced when the estimates were made while looking at photographs rather than at the real *S*.

When spectacles can so emphatically influence our judgments of personality traits, we may well wonder how many other inconsequential stimuli determine or aid in determining such judgments. Indeed, we adults must be bundles of biases which are at odds with the real make-up of the world. How do these biases arise? By learning, probably, but in most cases the conditions fostering that learning are obscure. We may guess that glasses are quite strongly associated with reading, that reading is associated with school work, that school work is associated with education, and education is associated with intelligence. Perhaps the advertisers of glasses have been partly responsible for this bias. Regardless of how it arose, it is clear that such a bias is present and that the objectivity of our judgments may suffer as a result.

Judgment of skin color among Negroes. We are well aware that Negroes differ in color. There are those with light skin and those with very dark skin. The question we are concerned with is how judgments of the amount of color of Negroes are affected by the color of a judge who is himself a Negro. The study which gives us this information was performed by Marks (247) at Fisk University, and the results show a definite bias which affects the judgments.

Marks had a group of students rate all other students in a class on several different factors. An 8-point scale was used with only the extremes of the dimension labelled for each factor. Six characteristics were rated, these being *energy*, *personal charm*, *intelligence*, *stoutness*, *skin color*, and *how well they knew* the person they were rating. By the rating method used, each person rated every other person on all traits, and each person was rated by every other person on all traits.

Before examining the data on the judgments of skin color as a function of the individual's own color, we need to examine a methodological problem which arose in this study. In order to get a measure of the deviation of the judgment of skin color from the true color we would have to have some measure of that true skin color. One measure which suggests itself is that of the average skin color assigned a given *S* by all other *S*s. Such an assumption would presume that personal biases would influence the judgments in such a way that these judgments would be equally distributed both above and below the true color, hence, would cancel. Marks made a check on this assumption by using a technique similar to the method of average error. The skin color of the inner surface of the upper arm was matched with a color wheel in which various components of red, black, yellow, and white could be fused. A color wheel is simply a rapidly rotating disk. If various colors are put on the wheel these colors will fuse when the disk rotates at high speed. The given objective measure for the matching of the skin color was the per cent of black paper which was necessary to make the match, with the amount of red, yellow, and white held constant. The more black paper required, the darker the skin. This percentage value was then correlated with the mean ratings of skin color and found to be .89, indicating a high degree of agreement between the two measurements. Marks concludes that he can use the average ratings as a satisfactory index of true skin color.

Our major interest concerns the ratings given other Negroes by a rater of a given skin color. The data show that a darker judge will rate another *S* lighter in color than will a light judge. That is, given a person of, say, average darkness, a very dark judge will assign this person a rating which is considerably lighter than that assigned by a very light judge. It appears that the reference point

around which the judgments of skin color revolve is the rater's conception of his own color. *S* is seen as lighter or darker than the rater. Relatively, then, the position of a person on the scale will be the same for all raters, though the absolute ratings will vary a great deal. In effect, each Negro has his own anchoring point (his own color) which determines his judgments of lightness and darkness in others. It may be pointed out, incidentally, that the preferred color among Marks' *S*s was neither a very light color nor a very dark color, but a color slightly past the middle of the scale toward the light end.

One other finding of Marks is of some interest. Correlations between average skin color ratings and personal charm were significantly high but negative. The correlations ranged from —.47 to —.82 for four different groups. The darker the skin the lower the rating for personal charm. There may be associated physical traits or unknown attitudes which cause the relationship, or the two may not be causally related at all. It may simply represent a halo effect in the ratings if the lighter skin color is preferred, though this is not likely, since, as we have seen, there is not a direct relationship between lightness of skin and preference.

Marks' findings can probably be reproduced on personal characteristics of different kinds in both white and colored people. Weight among girls, for example, would be an interesting characteristic to explore. How does a skinny college coed judge the weight of other coeds? How would a fat college coed's judgments compare? On any personal characteristic on which a high premium is placed in our society we are likely to find the egocentric biases entering in to modify the judgment. Each person ties his scale of values around his conception of the amount to which he possesses a given characteristic and judges others accordingly. We do, it appears, have an anchoring effect here in the same sense that we have an anchoring effect in the judgment of affective stimuli as discussed in the previous chapter. It would be an interesting experimental procedure to endeavor to change the anchoring values. Studies which we shall consider subsequently have attempted this in some areas.

A change in food preference by social example. We have previously considered an experiment in which food likes and dislikes were changed by a clear-cut learning situation. The present study

is concerned with change of food preferences through the social example of a respected leader. This study also gives rare data in that the stability of the changes has been studied.

The study was performed by Marinho (246) and was carried out with 4- to 6-year-old Brazilian children in the kindergarten of the Institute of Education at Rio de Janeiro. The initial problem was to determine the already existing food preferences of the children. The food, fruit pastes of different flavors, could be cut in pieces of exactly the same size (no mention is made of the color). The method of determining the favored and less-favored pastes was by paired comparisons, the various pairings taking place over a period of several days. From these initial measurements two experimental groups were segregated: (1) a group of 22 *S*s who had a strong preference for one of the foods used; and (2) a group of 7 *S*s who showed no preference for the foods. This latter group did not demonstrate even a temporary preference for one food over another.

In addition to the two experimental groups, two control groups were formed. Each *S* in the control group was matched with one in an experimental group on the basis of initial preference. During the experiment the control groups were subjected to no specific training of any kind; rather, they simply continued to choose between foods as they had done previously. Changes in food preferences by the control group showed how much modification could be attributed to normal living and eating conditions and to the continued choosing of foods. The difference between these changes and the changes shown by experimental groups gave an index of the influence of the experimental variable.

The training procedure was one of exposing a given *S* to a choice situation which was identical with the one he had just seen another child, a leader, experience. The leaders used had been picked by *E* after several weeks of observing the social relationships among the children on the playground. The several leaders selected were taken into *E*'s confidence. Before the leader and *S* came into the experimental room, *E* instructed the leader concerning which food he should choose.

The instructed leader and *S* came into the room where *E* was already sitting. *E* introduced the choice situation by means of a story with which the children were familiar. A Big Mouse (the

leader) and a Little Mouse (the S) were on their way to the Cat's house. When they reached the house the Cat didn't "happen" to be at home, so the two mice decided to eat the food in the Cat's house. The Big Mouse took a piece of food (we will call it Food B) from a plate on which both Food A and B were present. In the case of the experimental group in which there was a dominant preference for the food which we shall call A, the leader had been instructed to take Food B, i.e., the non-preferred food as far as S was concerned. In the case of the experimental group whose members had no preference for either Food A or B, the leader was instructed to take a given food and always take that food. We shall call this Food B also.

After the leader had chosen Food B, he said to S: "Now, Little Mouse, don't be afraid; take a nice bit of food from this other little plate." This other plate contained two pieces of food exactly the same as those contained on the plate from which the leader had made his "choice." After S had made his choice, E recorded it and the trial was over for that day.

If this "Big Mouse-Little Mouse" patter seems somewhat alien to the solemn "brass instruments" of the laboratory, it may be mentioned that in experimenting with children certain steps must be taken to insure their full coöperation. Asking these Ss to imagine that they are characters in a story is a good device for removing the strangeness of the experimental situation. Incidentally, it is not known whether the taking of food from the Cat's house made stealing acceptable, hence gave a push to the development of juvenile delinquency among the Ss.

We are not given information as to how many training trials Ss in the two groups had, but it appears that this was varied, some having many, others having but few. It will be remembered that the control group received none of these training trials. The results on the first ten training trials are outlined in Table 12. We have two experimental groups and their two control groups. During the training trials the leader always took Food B.

The upper half of Table 12 shows that changes were brought about as a function of the leader choosing the non-preferred food just prior to the time that S had to choose. The most preferred food at the start of the training was reduced in percentage preference and the least preferred food at the start of training was

increased in preference value. Some small changes were noted in the control group although these may well be within the range of the error of measurement. The change in the control group should be subtracted from the change in the experimental group to get the best estimate of the absolute amount of change which occurred as a result of the experimental manipulation.

TABLE 12

INFLUENCE OF LEADER PRESTIGE ON MODIFICATION OF TASTE PREFERENCES

Group	*Choice Before Training*		*Choice During Training*	
	% A	% B	% A	% B
Experimental I	87.3	12.7	58.6	41.4
Control I	89.6	10.4	84.6	15.4
Experimental II	52.9	47.1	11.4	88.6
Control II	51.4	48.6	42.9	57.1

Data from Marinho (246).

The lower half of the table shows that when no preference existed at the beginning, a preference could be rather firmly established. Food *B*, having been selected by the leader, was selected nearly 88 per cent of the time by *S*s in Experimental Group II. Again, the control group shows a small change in the same direction. All in all, these data point to the fact that the leader's behavior was a powerful factor in the choices made by *S*. Concerning the amount of change which took place in the two experimental groups, it would appear that it is easier to establish a new preference by this method than to break down an old preference and at the same time establish a new one.

The preferences of the children for the food were tested at various intervals for a period of a year. In the case of the food preferences set up in Group II (in which there was no preference for *A* over *B* at the start of the experiment), it was found that the preferences endured time after time even though the leader was no longer present. In the case of Group I this was not true, for with the passage of time there was a regression to their previous proportions of likes and dislikes. Only 17 per cent of the *S*s showed indications of enduring preferences for the foods.

This experiment demonstrates that a bias in one's preferences may be introduced by prestige suggestion; to be more specific, the food choices among leaders will determine to a certain extent the food preferences of followers in children's groups. The process might be thought of as imitation or suggestion. Neither term explains the facts but probably indicates that a complex form of social learning is taking place, a fact which we have suggested by the continual use of the term *training*.

Summary. In this section we have considered experiments in which the discriminal responses are clearly influenced by experimental conditions but in which the basic mechanism or mechanisms which bring about that response modification is obscure. In most cases it would appear that previous learning is the major factor, but because it is very difficult to analyze the complex of conditions, we have preferred for the time being to refer to the mechanisms as biases, the origins of which are unknown. It will be noted that we have not considered, in this section, experiments which are concerned with simple sensory discriminations. As we indicated before, biases which operate in those situations are called constant errors and in many cases special names, such as space error, time error, and so forth, have been applied. The causes for these errors may or may not be known but the errors do, as in the experiments reviewed in this section, produce a shift of response in a constant direction.

DESIGN METHODS I AND II: DIFFERENT GROUPS FOR EACH CONDITION

We need to evaluate certain design problems which have been inadequately discussed in reporting the experiments in this chapter. In this section we shall give the background for two design methods and in the following section attempt to demonstrate how these methods work out in actual practice.

Preliminary considerations. An experiment involves at least two different conditions. The purpose of the experiment is to determine the differences in response produced by these two conditions. In most of the experiments outlined thus far in the book, *S* has been tested or measured many times in the two or more conditions used in the experiment. Influences of different stimulus conditions

have thus been derived from a single *S*, or at best, a few *S*s. In most of the experiments to be reported in the remainder of the book, the estimated effect of a given condition is obtained by taking a single measurement on each of many *S*s. The mean performance of the group is taken as the best estimate of the influence of the condition. Performance under at least one other condition is obtained in a like fashion and the statistical difference between the means evaluated for the probability of significance.

The use of single measurements on many *S*s creates several important methodological problems, some of which we will consider in detail at this point. To simplify the exposition, let us assume that we wish to find the differences in performance produced by two conditions, *A* and *B*. There are two general ways of approaching the problem: (1) by using a different group of *S*s for each condition; or (2) by using all *S*s under both conditions. At this point we shall consider the two specific methods under the first general approach, i.e., cases in which a different group of *S*s is used in each condition. In a later chapter (X) we shall consider specific methods of the second general procedure.

When a different group of *S*s is used for each condition we may employ one of two general design schemes:

Design Method I: Random Groups
Design Method II: Matched Groups

Design Method I: random groups. By this method two separate samples of the population are taken, one group or sample serving under Condition *A*, the other under Condition *B*. (If there are more than two conditions additional samples will be formed, of course.) Each group is formed by drawing randomly from a single, defined population. Ideally, to accomplish this, we put the names of all members of the population into a big bowl, stir them, and draw out, say, 100 names of *S*s to work under Condition *A*, and 100 for Condition *B*. Usually such a procedure is quite impractical. In practice we would by some means get 200 *S*s from a defined population by a method that at least approximates random selection. The 200 *S*s would then be divided into two groups on a random basis. For example, the first *S* appearing at the laboratory would be put in the *A* group, the second in the *B* group, the third in *A*, and so on.

The basic assumption of Design Method I is that random assignment of Ss to groups will result in groups which do not differ significantly on any variable effecting the measured response. E runs one group under Condition *A*, the other under Condition *B*. If a significant difference between the mean performance of the two groups is found, the difference is said to be produced by differences in the *A* and *B* conditions. If no significant difference is found, *E* concludes that the two conditions do not differentially influence the response.

No telling argument can be advanced against Design Method I if the method of assigning *S*s to groups is truly random, and if the groups are large. Many experiments have been performed using this design. Yet, on the logic that other methods do not rest on the basic assumption of Design Method I, this method is not recommended if it is possible to use another. If we get a significant difference between the means when using Design Method I there is always the haunting possibility that this difference could be due to differences in the groups *as such* and *not* to differences produced by the experimental conditions. With large groups of *S*s and truly random selection, the risks of the method are slight, but when small groups are formed the risks are too great to warrant its use. There are studies reported in which groups were formed by random selection and then when both groups were measured on the same task significant differences in performance were found. The writer has often divided his class into groups by random selection of differently colored capsules which *S*s drew from a box. Occasionally this procedure has resulted in groups which differed in performance on the same task. No one can deny that we should get equivalent groups if our selection is random, and yet in actually forming small groups it does not always work out that way.

In summary, then, Design Method I is not recommended for small groups if it is possible to use one of the other methods. With large groups of *S*s and random selection, little risk is involved.

Design Method II: matched groups. This method has arisen largely to avoid the assumption upon which Design Method I is based. Rather than assume that random selection will result in two groups which are equal on pertinent variables, we know that this method assures equality by matching them on a pertinent

performance. There are two ways by which the matching can be accomplished.

1. *Matched on basis of means and variability.* *E* gives all *S*s practice on a given task and on the basis of these scores forms two groups so that the mean performance, and variability of performance of the two groups do not differ significantly.

2. *Matched pairs.* *E* gives all *S*s practice on a given task and on the basis of these scores pairs *S*s so that the score of one member of each pair is the same, or very nearly the same, as the other member of the pair. One *S* of each pair is put into one group to serve under one condition, the other *S* into another group to serve under a different condition. If more than two conditions are to be run, more than two *S*s will be matched at a time.

Either of the two methods of forming matched groups is satisfactory. To determine significance of difference between means, somewhat different statistical methods are used in each case for estimating the standard error of the mean difference. The first method (matching on means and variability) requires the calculation of a coefficient of correlation between performance on the matching task and the task used in the experiment. Since the calculation of a coefficient of correlation is somewhat laborious, and since the second method does not require it, the second method is preferred.

We need next to consider the nature of the pre-test by which we match the groups using either method. We have two kinds of pre-tests which have been used as matching tasks.

1. *Different but highly correlated tasks.* If we have a task, on which the performance scores are highly correlated with scores on the task to be used in the experiment, it may be used as a source for matching data. *The essential requirement is that the two tasks be highly correlated.*

Often, the beginning experimentalist will attempt to match groups on everything under the sun, including hair color and fraternity affiliation. Intelligence tests have probably been much overworked as criteria for matching without due consideration to the possible low correlation between these tests scores and scores on the experimental task. All we need is to be sure that the task from which we derive scores for matching is highly correlated with the task to be used in the experiment; *the precision of*

matching is directly related to the size of this correlation. No assumption concerning a correlation between two tasks should be made. One cannot observe the skills required to master the two tasks and because the skills required *seem* to be the same, assume that they are correlated. Too often an empirical determination of the correlation will show the relationship to be low. If the correlation between performance on the pre-test and performance on the experimental test is low, we are no better off than if we had selected random groups.

2. *Initial performance on experimental task.* It is often convenient and feasible to give all *S*s a few trials, or a short period of practice on the task to be used in the experiment. All *S*s get this practice under exactly the same conditions and then two groups are formed on the basis of these scores. Either matching procedure may be used. One group is then tested under Condition *A*, the other under Condition *B*. The correlation between performance on initial trials and experimental trials must, of course, be high to gain the precision desired. Since the two sets of scores are performance measures on the *same task*, the correlation is usually quite high. For this reason, this method is more efficient than matching on the basis of scores on a different task.

In certain instances this procedure is used in conjunction with Design Method I. Random groups are formed and all groups are given initial practice under exactly the same conditions. By the scores on this initial practice *E* can determine whether or not the random selection resulted in two groups which do not differ significantly on the task to be used under the experimental conditions. Occasionally in such instances it is discovered that random selection has not resulted in equal groups. If such is the case, *S*s can be shifted to equalize the groups.

It is our recommendation, then, that Design Method I be rejected if it is possible to use Design Method II. There are certain types of experiments which almost require Design Method II; hence, it becomes necessary that we have a thorough understanding of it. We shall later (Chapter X) point out the particular kinds of experiments for which the method is almost mandatory.

THE EFFECT OF LABELLING: AN EXERCISE IN EXPERIMENTAL DESIGN

In concluding this chapter we will evaluate three studies which show varying degrces of adherence to scientific procedure as defined by Design Methods I and II. As the subject matter we shall use studies which have been concerned with the influence of labelling on judgmental processes. Several experiments in the journals show that labelling a statement or an idea as emanating from a person of high repute will influence its acceptance by *S*. We shall consider one study which provides an illustration of a fairly decisive violation of scientific procedure; one which causes hesitation about accepting the results, and finally, one which is outstanding for rigor of design.

Use of the mails and random selection. The first study (52) suggests that we should be very cautious concerning the gathering of data through the use of the mails. A scale measuring religious attitudes was constructed. Initially the scale was mailed to 500 people of varying faiths. Out of the 500 sent, 213 (less than 50 per cent) were returned. The scale was sent out again to the 213 respondents, but three different conditions were imposed, each condition being given to approximately a third of the total. The conditions were as follows:

a. Expert opinion labelling: Respondents were shown how 20 clergymen had marked the attitude scale.

b. Majority opinion labelling: Respondents were shown how, on the average, the 213 people had marked the scale.

c. Control (no labelling): Merely a re-administration of the scale.

Out of the 213 scales sent out on this second run, 92 were returned, but only 65 were marked so that they could be used in the final analysis. The report does not indicate the exact number of cases in each of the three groups, saying only that they were about equally represented. We may presume, therefore, that there were about 20 cases in each. The results show that the expert opinion and the majority opinion conditions brought shifts in the attitudes which were greater than the shifts shown in the control group.

It seems clear, however, that such a study is inconclusive because of the high rate of loss of questionnaires. For all we can

tell, the people who did not return their questionnaires might have shown attitude shifts in the other direction *had they filled out the scale*. The fact that certain people did return the scales and other people did not indicates the presence of two populations which differ on at least one factor, namely, willingness to coöperate by returning the questionnaires. These populations may well differ in their reactions to the experimental variables, but this study provides no way of knowing whether this is true or not. Thus, there is no proof that the three groups, totalling 65 people, are representative of the original 500, nor that the three groups as such are representative of the same population. One cannot justifiably make comparisons between groups obtained by such methods. No bias as a result of labelling can be said to have been demonstrated as a consequence of the procedure used.

The use of random groups. In this experiment (25) *E* gave four different sections of elementary psychology students two passages to read and evaluate. One of these passages might be described as very radical in nature, the other strongly democratic. Naming the attitude of the passages isn't important for our purposes since almost all the students indicated that the democratic passage represented their point of view. The four sections of students were treated as follows:

Section 1. Read both passages and indicated which represented their point of view. No source for the statement was given the *Ss* by *E*.

Section 2. Same procedure as Section 1, but at the head of the first statement this phrase appeared: "Statement *A* is a fascist statement." Before Statement *B*, this phrase appeared: "Statement *B* is a Communist statement."

Section 3. Same as Section 1.

Section 4. Same procedure but Statement *A* was headed by: "Statement *A* is a Reactionary statement." Statement *B* was headed by: "Statement *B* is a Liberal statement."

After indicating which passage represented his point of view, *S* was to indicate how fully he agreed with the point of view. This was done on a 5-point scale. Thus, if *S* agreed with Statement *B*, a rating of zero would indicate bare acceptance; 1 would indicate somewhat more agreement, and so on, so that 5 would mean he fully agreed with the passage. As indicated previously, nearly 100 per cent of all *Ss* agreed with Statement *B* despite the

derogatory labels which were attached to it in the case of some of the sections. Differences arise, however, in the mean ratings of agreement in the different sections. The mean scale values of the ratings for each group were as follows:

Section 1: 4.0	Section 3: 4.0
Section 2: 3.6	Section 4: 4.3

We may note that the two control sections—the non-labelled groups, Sections 1 and 3, received identical mean values. The difference between the mean ratings of Section 2 (3.6) and Section 4 (4.3) is statistically very significant. This suggests that the derogatory label given Statement *B* for Section 2, while not necessarily prohibiting acceptance of the statement, did decrease the average ratings which indicated how much *S*s agreed with it. On the other hand, the allegedly laudatory label of "liberal" given Statement *B* for Section 4 enhanced the degree with which *S*s agreed.

We might conclude that while labelling in this instance does not change the basic opinions of the *S*s, it does decrease or increase the intensity of the opinion. Now, this is very likely the case, but can we, from this experiment, really conclude that? Do we know that the basic political "mean" and "variability" of Sections 2 and 4 were the same at the start? The answer, of course, is *no*. The obvious equality of Sections 1 and 3 does not thereby indicate that Sections 2 and 4 were equivalent at the beginning. Both groups had a relatively large number of *S*s in them, 73 and 60 respectively, and the groups were presumed to have been formed by random selection (Design Method I). However, it is noted that the standard deviation of the distribution of ratings for Section 4 is .68 whereas that for Section 2 is 1.3. Is this difference in scatter and the difference in the means a function of the experimental variable, i.e., the differences in labelling, or are they characteristics of the groups as such at the beginning? There is no way of telling from the data. We cannot, then, be completely sure that the difference obtained is a function of the variable introduced or of the groups before the introduction of the variable. We are likely to accept the first alternative, but if we are rigorous in our thinking, we will deny that we have clear-cut evidence for such a conclusion.

What was needed in the experiment above was some way of getting an indication of political beliefs before the experiment began. The ascertained amount of change (if any) would be demonstrated to be the result of the experimental procedure. It may not be an easy job to apply this experimental principle in some cases, but it is absolutely necessary if the claimed results are to be accepted without reservations.

Contrast the two above studies with the one to follow as regards the rigorousness of controls.

The use of matched groups. If certain statements, some labelled "rumor" and others labelled "facts," are given to *S*s who have no reason to suspect the truth of the labels, there should be differences in their willingness to accept the statements. If this is true, we can at once see how frequently bias may develop in our culture and how our judgmental processes may be swayed. A carefully done experiment by Smith (359) will indicate the influence of such biases in one restricted area.

Initially, Smith collected 100 statements about Russia, half of them supposedly favorable toward Russia, the other half unfavorable. Through a series of filterings 26 items were finally retained, 13 pro-Russian and 13 anti-Russian according to the opinions of several groups of judges. This final set of statements was believed to be plausible enough to permit college students to accept them as facts, but really there was little if any reason to believe that they were true. In essence, the statements were "fictitious facts." These 26 statements were randomized and mimeographed in three forms. One form bore explicit statement that all items were *actual facts* and had been verified as such by *E*. A second form carried the notation to the effect that these statements were *unverified rumors*, and that *E* had been unable to find facts to support them. The third form, used as a control, did not label the statements in any way.

The three forms were handed out at random, one each to a large group of elementary psychology students and each of these *S*s, regardless of the form he had, was to mark on a 4-point scale his degree of belief in each statement. A zero weight was given to the scoring if *S* checked (for a given statement) the phrase: "I do not believe the item to be true at all"; 1 point was given if *S* checked the phrase, "I wonder whether it could be true"; 2

points were given if *S* checked, "I believe it to be partly true; there may be something in it," and 3 points were given if *S* checked, "I believe the item to be true." *S*, of course, checked only one of the four alternatives for each statement.

In addition to checking the statements, *S* filled out an attitude scale which gave an index of his attitude toward Soviet Russia. This scale had been constructed previously, it had been standardized, and in an indirect fashion, validated (357, 358). Smith was thus able to equate his groups on the basis of their attitudes toward Russia, the most pertinent variable on which they could be equated for studying the influence of labelling on the statements about Russia. From the results of the attitude scale, three different attitude groups were picked, those pro-Russian, those neutral, and those anti-Russian. Each group had 90 *Ss*, 30 having been given the rumor-labelled form, 30 the fact-labelled, and 30 the no-label form. The summary of the nine groups, 30 to a group, is as follows:

Attitude	*Fact Label*	*No Label*	*Rumor Label*
Pro-Russian	30	30	30
Neutral	30	30	30
Anti-Russian	30	30	30

This design, in effect, yields two experimental variables: (1) the individual's attitude toward Russia; and (2) the type of label on the 26 factual statements about Russia. Each of these dimensions is sampled at 3 points. *E*'s basic working rule is that he should control all variables except one, and that one he varies. In terms of this rule, Smith has done three experiments, not individually, but at one time. The three experiments, each exploring a single dimension, are represented by any row or column. Thus, one experiment holds attitude constant, say, pro-Russian, and varies the labelling in three ways, and so forth.

Smith scored the anti-Soviet statements and the pro-Soviet statements separately so that the maximum score for *S* on the total of 13 statements in each section would be 39. If for each statement *S* checked the phrase, "I believe the item to be true," he would get 3 times 13 or a score of 39. The mean scores for the 30 *Ss* in each of the nine groups for the anti- and pro-Russian statements are shown in Table 13. The separate scoring of the

anti- and pro- statements actually makes a third variable in the experiment, but with only two points of the dimension sampled.

TABLE 13

EFFECT OF LABELLING ON ACCEPTANCE OF STATEMENTS ABOUT SOVIET RUSSIA

	Fact Label		*No Label*		*Rumor Label*	
Attitude	Pro	Anti	Pro	Anti	Pro	Anti
Pro-Soviet	26.23	17.13	19.37	10.07	20.17	9.50
Neutral	21.60	18.47	17.07	13.80	15.77	13.00
Anti-Soviet	16.60	21.13	14.80	16.33	13.97	15.57

Data from Smith (359).

From the table we can see that, in all cases, *S* is more credulous towards the statements labelled factual. There are no reversals to this tendency. Differences in acceptance or rejectance of the statements are also evident as a function of *S*'s attitude toward Russia. The more favorable the attitude toward Russia the more likely *S* is to indicate belief in the statements which are favorable toward Russia and the less likely he is to indicate belief in the statements which are anti-Russian. Again, there are no exceptions to these trends. This experiment should be studied carefully, since it is an exceptionally fine illustration of careful dimensional analysis. And, the results show once again that conditions bias our responses, even though the experiment shows nothing about the origin of the biases.

CONCLUDING CONSIDERATIONS

In this chapter we have examined experiments which demonstrated how (1) learning may influence simple and complex discriminal processes; (2) motivation may influence these same processes, and (3) other unanalyzed biases produce deviations in the discriminal processes. In part, the intent of this chapter has been to make you learning and motivation "conscious," for a great deal of the remainder of this book will be spent on the discussion of the experimental manipulations which cause these two basic behavioral determinants to vary in influence. It has also been the purpose of this chapter to make you aware of the fact

that a mere correlation of the physical stimulus dimension with the response dimension does not provide a complete picture of the causal factors in behavior. Even with the most simple sensory processes, subtle internal variables may be influential.

Our discussion of methodological problems in this chapter has centered around two basic Design Methods, I and II. These are two of the four basic methods which may be used in designing experiments. We shall have many occasions to refer to these methods by number, so the basic idea of each should be thoroughly assimilated.

CHAPTER VI

Motivation

INTRODUCTION

With this topic we continue to trace the sequence of events represented in Fig. 3, Chapter I. All areas of study represented in Fig. 3 from motivation onward are so closely related, their underlying processes so interdependent, that it has already been difficult to consider any one of them without reference to the others. Motivation and learning, for example, dovetail with each other so firmly that we have a difficult time separating them even on a verbal level.

The fundamental observation made concerning organisms is that they perform acts with respect to objects and events (real or symbolic) in their environment. A man boards a street car; a lady powders her nose; a gambler glares at the dice; a student registers for a course; a boy solves an algebraic equation; a rat builds a nest. Two basic facts are illustrated by these observations: (1) the organism is stimulated in some manner to do these things, and (2) it has the skills or abilities with which to do them. The first is a problem of motivation and the second is a problem of learning. Yet it is never as simple as that. For example, where did the motivation to powder the nose originate? It is absurd to work on the premise that a lady is born with an inherent tendency to powder her nose. Where did the motivation to solve an algebraic equation arise? Are we born with some slowly maturing desire to solve equations? It seems likely that the motive to powder the nose and the motive to solve equations have been almost entirely learned in the same sense that the *movements* needed to powder the nose and the *symbolic skill* needed to solve equations have been learned. But if we assume, as we shall, that motivation

is neccssary for learning to take place, we have a problem in determining how the motivation arose which allowed the acquisition of a motive to powder the nose. This really isn't double talk. The two processes, motivation and learning, turn upon each other; they are dependent upon each other if the organism is to maintain itself in the environment. We are going to separate the two for study purposes only.

DEFINITIONAL PROBLEMS

Motivation as a construct. The term *motivation* (or *motive*) is a logical scientific construct.* By this we mean that in its ordinary sense it is not an entity, object, or thing such as a living cell, an eye, or a piece of coal. It has no immediate sensory referent. It is an hypothetical process or state of the organism, the operation of which we infer from observations of the behavior of the organism. *Threshold* is a construct, though in our previous usage of the term we did not indicate specifically that it was. Motivation is a concept like *threshold*, like *gravity*, or like *intelligence*. You have never seen a threshold. Nor have you ever seen any gravity or intelligence. The term *intelligence* is used to refer to a certain dimension of behavior manifested in different amounts by different people as measured by the number of items which are correctly answered on a test called an intelligence test. In the same manner, we manipulate a certain class of experimental (stimulus) variables, observe a difference in the responses as a consequence, and infer that we have produced changes in what is called the motivation of the organism. We must, of course, specify the nature of the responses we are to observe as well as the characteristics of the class of stimulus variables which are manipulated. When this is done we have anchored our construct to both the stimulus and response variables.

Response dimensions related to changes in motivation. We have said that to use motivation as a scientific concept we must specify the nature of the response changes which are associated with motivational changes. We have two response dimensions from which we infer changes in motivation: (1) changes in performance; and (2) changes in muscular tension or energy output.

* For our purposes, *concept* and *construct* may be used interchangeably.

The most common behavior from which we infer changes in motivation is a change in performance on standardized tasks. If learning speed is increased we may suspect the motivation has increased; if work output goes up we may suspect that motivation has increased; if output goes down, or if learning is retarded, we may suspect that there has been a decrease in motivation. Note that we may *suspect* that these response changes are a function of motivation. *Change in performance is an unreliable criterion for inferring changes in motivation,* because we may observe changes in performance which are due to stimulus variables *not* in the class of motivational variables. For example, distributed practice (allowing rest periods) will usually enhance the speed of learning (change in performance), but little credence is given the idea that this is exclusively a function of change in motivation. In short, *whether or not change in performance is to signify a change in motivation depends upon the particular class of stimulus variables being manipulated.* The latter part of this chapter will be devoted to outlining these variables which are believed to be associated with changes in motivation.

The most meaningful response change from which we infer motivation is the amount of energy output or muscular tension exhibited by the organism. This index of change in motivation is used less frequently than the first, largely because of difficulties attending its quantification. Yet, change in activity level is the basic observation associated with motivational changes. The strongly motivated individual is muscularly active or at least muscularly tense; the weakly motivated individual shows signs of ennui, boredom, laziness, or lethargy. When you make an observation that an individual is highly motivated, you usually do it on the basis of the activity level of that individual.

Within certain limits, the two response changes indicating changes in motivation are correlated, but there is not a one to one relationship. If these two measures were perfectly correlated there would be no need for using both. We would work with the one which was least difficult to use in the experimental situation. The fact that they are not perfectly correlated raises some rather difficult problems. How can both performance change and energy change be used as measures of the same process when they are not perfectly correlated? Let us take an illustration. If we

announce to an assembly worker that we will give a bonus for every unit produced beyond a certain fixed quantity, and if his production increases following the announcement, we assume that there has been an increase in motivation. We may, at the same time, get indices of increased energy being expended as the production goes up. But, suppose there is an increase in energy output and not a corresponding increase in performance? Or, suppose there is an actual inhibition (decrement) of performance accompanying the increased energy output? Are we to say that motivation has increased in spite of the fact that performance decreased? We would have, in such instances, two independent measures of motivation which are not perfectly correlated. Which one is *the* measure of motivation? Neither, of course. We cannot speak of *the* measure of motivation when two measures have been used which are not perfectly correlated in the experimental situation.

It is our belief that we should reserve the term *motivation* for changes in energy output. There are two reasons for this: (1) it would give a clear-cut differentiation between changes in performance due to variables associated with learning and those associated with motivation; and (2) there would be less confusion than is the case when two poorly correlated response measures are used to infer the same process. However, it is impractical at present to adopt such a criterion for inferring motivation. Many of the data pertaining to motivational changes which are available are based on performance changes. Furthermore, there are many situations in which it is almost impossible to measure the changes in energy level. For the present, then, we cannot use energy changes as the sole index of motivation: we must also use the changes in performance and at the same time recognize that the two are not perfectly correlated.

Why use the term? If you examine the above situation, you might infer that it would not change the nature of the empirical facts obtained by experimentation if we completely eliminated the term *motivation*. For example, when we discover by sound experimental procedure that output goes up as a bonus is given, that stands as a scientific relationship. Why go any further? Why say that increased output indicates increased motivation? The reason for the use of the term is a scientifically practical one.

Without such a term we have a mass of unrelated facts on our hands. However, if we find that a considerable number of experimental operations will produce a change in performance (or energy output), and if it appears that these stimulus changes are all affecting the same basic process, then our knowledge is integrated by the use of a single term for that process. In this case the term is *motivation.*

Constructs such as this allow for more generalized forms of explanation. Explaining observed phenomena on the basis of the immediate change in the stimulus variable, is, as we indicated in Chapter I, a low level form of explanation. Such a procedure does not allow for generalized explanatory concepts. Each fact in and of itself would have to be explained by the change in the stimulus conditions. When, however, there is believed to be communality in the processes produced by those stimulus changes, our explanation becomes more generalized by referring to those processes with a single term. Without such a generalized concept it would be impossible logically to predict behavior in a new situation. We could only state that such and such *had* happened under such and such highly particular circumstances. Fortunately the scientists isn't content with the level of particulars. He reaches out for the *general* concept, the inclusive formula, the law. So when there is a common basis in experimental manipulations of stimulus variables, and when these manipulations produce similar changes in response, the use of a construct permits explanation at a more general level than does reference to each of the operations individually.

On the conceptual level we may think of motivation as a state which energizes the organism. This state is in turn produced and varied by manipulation of environmental conditions by *E.* In animal studies we can give, in most cases, a precise quantitative definition of motivation in terms of the number of hours of privation of food, water, mate, and so forth. Such operational definitions are not easily formulated for human *Ss.* In human adult experiments our definition hinges on the instructions we give *S.* In these instructions we may indicate rewards and punishments which *S* may get, or we may try to vary certain aspects of motivation in terms of instructions alone, without threat of punishment or enticement of reward.

Strictly speaking, we shall have an operational definition of motivation for each experiment, but we shall find that most of them can be classified under a few general headings. The classes of operations used to produce and measure changes in performance and changes in energy output must remain our operational definition of motivation.

Terminology troubles. The study of motivation is complicated by the multiplicity of terms which have been used by various writers in referring to what seems to be the same aspect of behavior. There have been instincts, drives, wants, needs, propensities, appetites, interests, aversions, desires, values, sets, attitudes, valences, forces, and others. All of these terms refer to certain facets of motivation. In most cases each author has defined his own terms for his own *writing* (which is proper and necessary), but students have had to suffer from a multiplicity of both concepts and vocabularies. We shall limit the number of concepts as much as possible.

PRIMARY AND DERIVED MOTIVES

Primary motives. There are certain primary conditions which arouse behavior. Among these are conditions which give rise to thirst, hunger, fatigue, sexual needs, and pain avoidance. It is these primary motives which are employed almost exclusively in the animal laboratory. The strength of these motives may be defined in terms of hours of privation in the case of hunger, thirst, and sex and in terms of hours without rest, for fatigue. Pain avoidance is very difficult to manipulate quantitatively.

While we recognize similar basic motivating conditions in human beings, we have not found it possible, in general, to vary them in experiments with human subjects. In laboratory experiments, the use of electric shock seems to be the only major method of arousing primary motives in humans, and there is even some question as to whether its influence is a direct result of sheer pain. We do not deprive an adult of food for 24 hours and then offer a filet mignon if he performs a given task at a certain speed, although such a procedure might well be worth trying. The fact is that experiments on motivation with human adults aim at the study of what we choose to call derived motives.

Derived motives. These motives are so named because unlike primary motives they are believed to be largely learned. These motives have been variously named, have varied greatly in number (in part because of different descriptive dimensions used), and are still subject to considerable controversy, largely because the operations which evoke them are quite obscure. We may point out some of the *descriptive terms* which have been used to identify these complex motives; motive to preserve self-esteem, to gain prestige, to dominate others, to be gregarious, to achieve one's deepest ambitions, or, for that matter, the motive to see a ball game. There is nothing sacred about any such list. For the present we need to emphasize only that there *are* complex motives which rather completely prompt adult behavior.

DEVELOPMENT OF MOTIVE-INCENTIVE RELATIONSHIPS

In dealing with motivation we are concerned primarily with its influence on performance phenomena associated with learning. We ask the question: "What rôles does motivation play in learning?" In answering this question we shall try to avoid certain problems which are controversial in nature and delay treatment of the difficult questions until we have finished the study of learning later in the book. In the current chapter we will outline events and interrelationships which we think can be accepted with least hesitation.

Following Melton (254), we may recognize three basic rôles which motivation plays in the learning process: (1) the energizing rôle, (2) the directing rôle, and (3) the emphasizing or selecting rôle.

The Energizing Rôle

As suggested in Fig. 3, Chapter I, the motivated organism, when faced with a problem situation—a situation in which a "correct" response is not immediately available—will show variable behavior. The amount of variable behavior or general activity is a direct index of the strength of motivation. If the situation is one in which the organism must discover the correct response, it is obvious that the greater the number of responses tried, the greater the probabilities of hitting upon the correct response.

It would appear, then, that increased energy level may be an important factor in learning just because of the increased likelihood that the correct response will be found. We shall not consider the relationship between energy level and speed of learning at this point; rather, we shall examine the methods which have been used to relate changes in energy output to certain changes in the stimulus conditions.

Animals. The procedure for animal studies is straightforward—deprive an animal of food or water for varying amounts of time and measure the changes in activity. A rotating circular cage mounted on an horizontal axle is one device used to measure activity level. A counter is attached to the axle so that the number of revolutions of the cage can be obtained for any desired unit of time. In the cage, as in a treadmill, the animal may run a great distance without getting anywhere.

Another device is the *spring* or *tambour-mounted* cage. Tambours are air tubes with a mouth covered with rubber sheeting. The tubes, joining together from the four legs of the cage, lead to sensitive markers. These markers write on a *kymograph,* a slowly rotating drum covered with waxed or smoked paper. The rat (or other animal), when active, bounces the cage on the tambours, thus changing the air pressure which in turn activates the markers which record the animal's activity.

Still another mechanism is the *tilting cage,* in which slight tilts of the cage (produced by movements of the animal) activate a mechanical or electrical counter (32).

With such devices it has been shown that in general the activity level is related directly to the hours of deprivation of food, water, or mate, these being the three conditions usually studied. Concerning the sex motive, it has been shown that the activity of a female rat is at a maximum every four or five days, (the oestrous cycle), at which time the rat actively seeks a mate (398). Concerning hunger, it is known that the greatest activity of a rat occurs just before eating and the least just after eating (311). If a rat is deprived of food for several days there is an increase in the amount of activity up to about 96 hours, after which time a gradual reduction (because of weakening) takes place (424). These results are enough to illustrate the point; within limits, as hours of deprivation increase the energy output increases, and we say that

this reflects an increase in motivation. The amount of activity is, of course, influenced by many factors such as age, heredity, nutritional balance, and drugs (305).

Humans. It is not practical to study human subjects in activity cages, though this technique has been used for other purposes in the case of newborn infants (180). Instead, we must infer from certain findings that the activity level of humans is associated with certain deprivations or unsatisfied states. Our inference relates instructions to perform to changes in energy output. We do not deprive the person of food and measure activity level; rather, we instruct him to learn a task and then measure the difference in tension or energy output before learning (comparative rest) and during learning.

The methods that have been used to measure tension differences between learning and rest have varied, and cannot be said to be highly correlated. Stroud (369) constructed a *hand stylus* so that it would measure air pressure changes produced by the grip of the hand. The stylus was a pencil-like cylinder used to trace the indented path of a maze. Differences in the pressure throughout the stages of learning, or as between one *S* and another, were measured directly.

Reflex responses, such as the knee jerk, may be used as indices of tension. A positive relationship has been demonstrated between the amplitude of the knee jerk and the amount of tension exerted on a *hand dynamometer* (a device for recording grip strength) so that the amplitude of the knee jerk at various times during learning would be indicative of tension changes (68). The reflex eyelid response (blinking) has been suggested as a measure of tension, with the suggested relationship being: the greater the number of blinks per unit of time the greater the tension. However, results are somewhat contradictory concerning this measure. There is some indication that heart rate is a possible index of tension (27) as is respiration rate also (375), but confirming evidence is needed.

The *galvanic skin response*, with which we are already familiar, has also been used as a measure of tension. It can be shown that the resistance of the skin is inversely related to the amount of tension, but the possibility that factors other than tension may bring about the same change in resistance has also been pointed

out. Consequently, we cannot accept without qualification the relationship of the resistance of the skin to tension level (78).

Probably the most satisfactory measure of tension is obtained by direct measurements of electrical changes in the muscles, i.e., the *action potentials* which result from contraction of the muscle fibers. With the great improvements which have been made in amplifying minute electrical changes, this method is probably less ambiguous than any of the others. However, it is not as yet adaptable to certain tasks where considerable overt movement is involved.

Each of the above is presumed to be an index of activity reflecting different energy levels. Since the correlations between the different methods are not perfect, we actually have as many different measures of energy or tension level as we do methods of measurement. These discrepancies may disappear with further research and standardized recording devices. (More detailed accounts of various methods used in measuring tension may be found in Davis, 78.)

With regard to the energy output changes *during* learning, we can make no general statements. Almost all combinations of increases and decreases have been found by various investigators, using various tasks and various methods of measurements. The most common finding has been high initial tension followed by a gradual decrease as learning progressed. Without exception, however, tension is greater during learning than it is during comparative rest. We infer that these changes indicate a state of increased motivation brought on by the attempt to comply with the instructions to learn or to perform a given task.

No E has yet made a serious attempt to dimensionalize tension by using judges. It is quite possible that scaling methods would result in reliable measurement, and that certain characteristics of behavior would be found to vary concomitantly with changes in tension as measured by mechanical devices. Most of our personal estimates of amount of motivation are inferred from such characteristics. Any instructor who has observed his students while taking examinations can name several responses which seem to increase in frequency and intensity; these responses are such characteristics as the flexing of the ankle of the crossed leg, finger chewing, and general exaggeration of tics or twitches

of one kind or another. There seems to be no reason why such behavior could not be quantified by judges rating *S* as he is put through a sequence of experimental conditions.

The Directing Rôle

Incentives. We seldom deal with organisms which are just plain motivated, without being motivated *toward* something. When an organism is motivated or energized, its behavior has directive aspects. A hungry rat is active until food is found and the searching behavior is likely to be directed toward places where food had been found previously. A student, thirsting for knowledge, is almost certain to go to the library and not to the fraternity house. Because of our past experience, we go to restaurants rather than filling stations when we are hungry. The general principle seems to be that we tend, when motivated, to make responses toward a goal which in the past has eliminated or reduced the strength of the motivational state. Learning has established a relationship between the motivating state and the goal-objects which reduce that motivational state. Thus, certain objects (real or symbolic), or classes of objects or states, are perceived as being satisfying to a given motivational condition. Such goal objects are called *incentives.* In studying the influence of motivation on performance we must usually include an incentive as part of the motivation complex.

In the case of animals the motive and the incentive are fairly clear-cut. A rat is deprived of food for a period of time and the food is placed at the end of a maze as an incentive. In the case of human *S*s the incentives of laboratory experiments are as obscure as the motives. What, for example, is the motive-incentive relationship when a student volunteers to serve in an experiment? All we can say is that the student has agreed to coöperate. What is the incentive? What is it that the student "gets out of the experiment" which brings about satisfaction in the sense that food reduces the motivation of a hungry rat? If he has a motive to coöperate, then coöperation is the satisfying incentive. This, however, is just so much naming, and scientifically is quite unsatisfying. Because of this unsatisfying state of affairs, psychologists are motivated to know more about these problems. We might, of course, ask *S* what the motive was that prompted

him to volunteer for the experiment, and we might ask him what provides the source of satisfaction, if the latter is apparent. However, clinical psychologists and psychiatrists have taught us to be skeptical of the reasons people give us for their behavior, since their underlying motives may be unconscious or unreported.

The facts of the case are that we know very little about the motive-incentive relationships in *S*s upon whom the overwhelming proportion of our data has been collected. This will come out more clearly when we start to examine in detail the conditions which *E*s have manipulated in an attempt to vary motivation. In most human experiments the directive function is supplied in part by the instructions to *S* along with the past experience of *S* with similar problems.

Illustrations of incentives considered thus far have all been *positive* in nature. In everyday life, as well as in the laboratory, learning is thought to be influenced by the administration of punishment or *negative incentives.* It is an "If-you-don't-do-this-you-will-get-this" type of situation. Electric shock has been used most often in the laboratory as a form of negative incentive. The influence of shock, as well as other negative incentives, may be both directive and energizing in function.

With human *S*s especially, there are many instances in which an incentive itself appears to have the power to energize as well as direct behavior. Most of our complex psychological motives appear to be long-term in nature and are relatively non-cyclical as is the case with hunger or thirst. Let us assume that *S* has a long-term motive to make money. He is sitting, relaxed in an arm-chair listening to the radio, when a friend enters the room and offers him ten dollars if he will write a term paper for him. With the exhibition of the ten dollars, *S*'s behavior changes markedly. He is now highly energized or activated to start work. It is as if the presentation of the ten dollar incentive "set off" or tapped a motive; or, it is as if a motive were "waiting around" for an incentive. It appears that the receipt of money has served to satisfy a motive in the past, a motive which we have probably simplified by calling it a motive to make money. Therefore the incentive is now perceived as being potentially satisfying to the motive which must have been present in some sort of latent and long-lasting form. Thus, incentives may not only provide the basis for the

acquisition of direction to motivated behavior, but may "set off" motives also.

What about the evidence for the directive function of motivation as it affects learning? We shall consider only animal data at this point since the directive function in the case of humans will be considered later in the chapter under the special topic of *set*.

Experimental demonstration of the directive nature of motivation. Animal experimentation is somewhat more difficult than it might seem at first. One doesn't *just* drop a rat in a maze and record the time it takes the rat to reach the food at the end of the maze. Many preliminary steps and controls are necessary. We shall follow a study through in some detail in order to understand some of the problems faced by the animal psychologist.

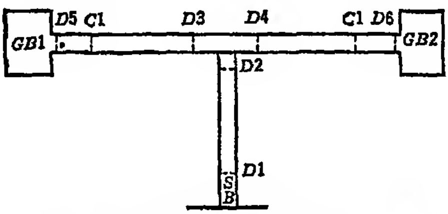


FIG. 26. Simple T-maze. This particular floor plan was used by Kendler (193). *SB* indicates starting box; *D1*, *D2*, etc., doors; *C1* and *C2* curtains, and *GB1* and *GB2*, goal boxes.

An experiment by Kendler (193) was designed to investigate the directive function of physiologically motivating stimuli. Rats were trained in a simple *T*-maze as shown in Fig. 26, under two motivating conditions, hunger and thirst, each defined by hours of deprivation. During training, food was found at one end of the *T*, water at the other. On test trials the rats were made *either* hungry or thirsty, one or the other of the motive conditions having been satiated. The question is whether or not the rat, after training under both motivating conditions, will go to the appropriate end of the *T* when either hungry or thirsty, e.g., if thirsty, will he go to the end where he has previously found water?

Preliminary training of naïve rats usually involves two steps. First, the rats must become adjusted to being handled. Kendler did this by handling each of his 20 rats for several minutes each day for a week. The second step involves adjustment to the

apparatus. In this experiment, two days were allowed for the second objective. On the first day the rats explored the maze for an hour. None of the doors was closed and neither food nor water was available in the food boxes. On the second day, *E* gave each rat two trial runs by placing the animal in the starting box and, as the rat moved down the alley, closing the doors behind him to prevent retracing.

Rats may have biases for making left or right turns. Kendler recorded the turns made by each rat on the preliminary trials and discarded any rat which consistently made the same turn at the junction. Moreover, during the regular training series, *E* forced half the rats to go to the left for food and to the right for water, and reversed the positions of the incentives for the other half.

All rats were under 21-hour deprivation at the time of the training series, and were believed to have been satiated nearly equally on water and food before the 21-hour period started.

> The training series consisted of four trials daily for seven days. The initial run of each day was a free choice, the second was a forced trial to the side opposite that chosen on trial one, the third was a free choice, while the fourth was a forced run to the side opposite that chosen on trial three. Thus, the animals had equal experience with the contents of both goal boxes. Forced runs were accomplished by lowering Door Three or Door Four (193, p. 215).

Test trials consisted of four runs, each run being given on a different day.

> On the first and fourth days of the test series the *Ss* were motivated for one goal object (food or water), while on the second and third day they were motivated for the other goal object. This *abba* order was utilized to avoid any possibility of building up an alternation pattern of responding on successive test trials.
>
> Since many animals had acquired preferences for one of the two alternative sides, the last four free choices on the sixth and seventh days of training were noted, and half the animals were motivated, on the first test trial, against the side for which they had a preference, while for the other half the opposite was true. Thus, on any one test trial the preference factor was approximately equated (193, p. 215).

On the four test trials the rats chose the appropriate alley 98 per cent of the time when motivated for water, and 73 per

cent of the time when motivated for food. The two groups combined gave 85 per cent correct choice. Chance would be 50 per cent. It is clear from these results that the animal's behavior is directed on the basis of internal motivating stimuli. It is presumed that this directive function is acquired as a consequence of the learning trials, where the rats were rewarded by incentives appropriate to only one of the two motivating conditions present at the time.

The directive function of motivating stimuli may be inferred from certain other animal data. Rats prefer bran mash to sunflower seed as a reward or incentive. Elliott (88) gave bran mash as a reward while rats were learning a maze until the tenth trial, at which time he substituted sunflower seed. On the eleventh trial the number of errors made in running the maze greatly increased. It was as if the rats were running to a specific incentive and the change in the incentive disrupted their performance.

The Emphasizing or Selecting Rôle

We have seen that following the awarding of incentives for performing a certain task, the behavior sequence preceding the reward tends to be strengthened—learning has taken place. When given precise statement this is known as the *law of effect*, which we shall accept as established for the present. The principle assumes that reduction in motivation, through interaction with an incentive appropriate to the motivating condition, tends to strengthen or select the responses which brought about this state of affairs. Indeed, the emphasizing or selective function of motivation develops the directive function, since it is only by such strengthening of responses that the organism's behavior shows less variability and greater efficiency as successive trials are given. Furthermore, it can be seen, the receipt of an incentive which reduces the motivating condition must necessarily affect the energy level of the organism, within certain limits. Thus, all three rôles are closely intertwined.

Symbolic incentive or reward. We have said that the incentive or reward received at the end of a behavior sequence will tend to strengthen or reinforce that sequence. If a rat is allowed to eat food in the goal box at the end of the maze after each run, the number of errors made, and the time required to run the maze,

are gradually reduced. Now let us inquire into the situations in which incentives are symbolic in nature, i.e., when they do not have direct reinforcing power as such. For example, how does it happen that money is avidly received as an incentive? We can't eat money, we can't drink money—it has no power to satisfy primary motives directly. If we assume, as most psychologists do, that our derived motives are built on primary motives as a consequence of learning, any information we can get concerning the origin of derived motives and the incentives which touch them off and satisfy them, will add to our knowledge of the motivation of human adults. We shall examine several animal studies to show a few basic mechanisms by the operation of which it is probable that our complex motive-incentive conditions are derived.

1. *Token experiments.* Studies by Wolfe (418) and Cowles (71) with chimpanzees have shown that tokens—poker chips—can become incentives and acquire reinforcing power. In the beginning the chimpanzees were taught to insert the chips into a slot machine to get grapes, which were considered to have direct or primary incentive value. The animals learned to differentiate between colors of the chips so that a certain color was found to be worth 1 grape, another 5 grapes, and so forth. Then the chimpanzees were put to work, doing manual tasks in some instances, for which they were given poker chips periodically, but they were not allowed to cash them in immediately for grapes. The animals might continue to work for chips alone for as long as 20 minutes before they would leave the work to try to cash the chips for primary incentives (grapes). Furthermore, not only would the animal work for tokens, but it would learn new tasks also, with only the chips serving as incentives. Clearly, the chips had gained incentive value for the animals. The analogy with human behavior is obvious.

2. *"Click" experiments.* Experiments by Skinner (356) and Bugelski (46) have demonstrated symbolic incentives in rats. The device used in these experiments has become standard equipment in one form or another and has come to be known as the *Skinner-box* (Fig. 27). We shall refer to it frequently throughout this book. A small bar projects from the end of a box. Immediately below the bar is a small food box (or water orifice). Pushing

the lever causes a small pellet of food to drop into the box. The rat, when initially placed in such a box, explores the interior and sooner or later hits the bar and releases a pellet of food. In order to speed the initial part of learning, some *E*s have put a small bit of moist food on the bar, or in some cases have blown the

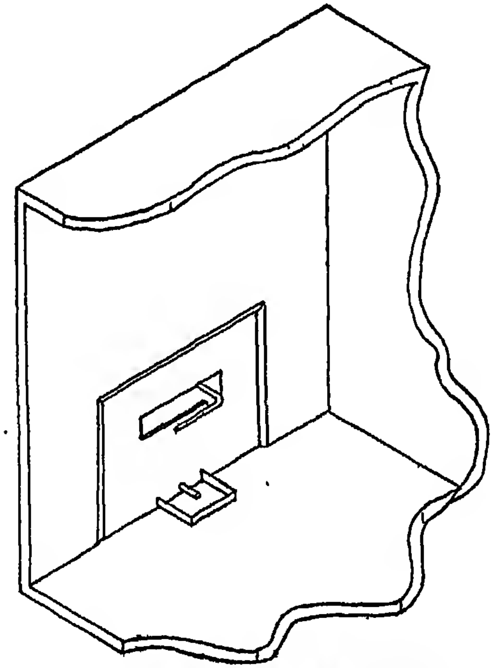


FIG. 27. Cut-away diagram of a Skinner-box.

odor of food through the slot from which the bar projects. These stimuli will cause the rat to concentrate its early explorations in the general area of the bar, thus increasing the probability of hitting it. Several days of habituation to the box may be necessary before the experiment proper begins. Learning may be measured by the number of bar depressions per unit of time, e.g., a minute. The pellets of food are quite small, so that it takes many of them to satiate a hungry rat.

In the Bugelski experiment, two groups of rats were used. One group was trained to depress the bar and get food. The second

or experimental group was also trained in the same manner but in addition to receiving a pellet of food each time the bar was depressed, a distinct "click" sounded. After the response was fairly well learned, both groups went through a process during which no pellet of food appeared when the bar was pressed. The removal of the food incentive (or the particular incentive being used for a series of trials) is called *experimental extinction.* During extinction neither group received food, but the experimental group continued to hear the click each time the bar was depressed. The rats in both groups were allowed to continue pressing the bar until such responses ceased (extinguished). The group receiving the "click" stimulus took 30 per cent longer to extinguish than did the control group which did not have the click. The click stimulus had gained certain incentive value in and of itself.

3. *The Brogden experiment.* An experiment by Brogden (36) illustrates a special case in the developing of a symbolic incentive and will introduce us to a new technique. Brogden worked with the flexion response in the dog. A small electrode used to administer shock was attached to the forefoot. A bell was sounded and the shock administered shortly thereafter. If the dog lifted his foot at the sound of the bell, he would avoid the shock. Needless to say, the dog soon learned to lift his foot when the bell sounded and thus did avoid the shock.

In this experiment, after the response was developed so that it occurred with nearly 100 per cent frequency to the sound of the bell, Brogden started giving the dog biscuits after each response and discontinued the shock. The training sessions followed 20 hours of food deprivation, so that the transfer in the nature of the reinforcement from punishment to reward took place without difficulty.

After the response was evoked at nearly 100 per cent frequency to the sound of the bell with food as incentive, the group of 12 dogs was split into three groups of four each. Group I was satiated with biscuit, placed in the experimental situation where the bell was rung, and food presented as usual if the flexion response occurred. The dogs in this group, being satiated, *did not eat the food.* Group II was treated the same as Group I except that *no* food was presented after the response. Group III

was *hungry* when placed in the experimental situation and *no* food was given after the response.

The results of the experiment are shown in Fig. 28 in terms of the per cent frequency of flexion response elicited on test trials. Ten test periods were given, 20 trials per period, each period on successive days. Each trial consisted of ringing the bell once. Figure 28 shows that Group I had only slight decrement in response over the 200 trials. At the end of this period, 73 per

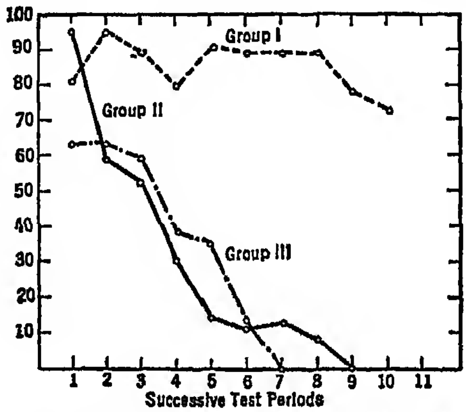


FIG. 28. The operation of symbolic reward in the maintenance of the flexion response in the dog. See text for explanation of different groups. Data from Brogden (36).

cent of the bell presentations were still being followed by flexion responses even though no food was ingested. Both of the other groups showed rapid extinction, i.e., the response died out so that it no longer occurred to the bell. Since Group II was not presented food following the response and Group I was, this variation in treatment must account for the differences in frequency of response. It would appear that the *sight* and *smell* of food have incentive properties so that the response continues to be evoked. Thus we have another kind of symbolic reward in which the stimulus which normally serves to reduce the strength of motivation, and hence bring about learning, will continue to perpetuate learned behavior even though the dog cannot be said to be getting direct motive reduction. This type of symbolic incentive needs to be produced in a situation where a new

response has to be acquired, although setting up such a situation would be difficult.

There are many other illustrations of the development of symbolic rewards in the literature, but these are enough to illustrate the point—the emphasizing function of incentives need not come from objects with direct physiological value.

From the considerations in this section we must grasp, in addition to the understanding of the various methods of experimentation, the three following systematic relationships:

1. *Primary reinforcement.* Achievement of an incentive appropriate to the motivation serves to select or emphasize the behavior sequence which precedes the obtaining of the incentive.

2. *Secondary reinforcement.* If other stimuli occur consistently with an incentive which is appropriate for the reduction of a motivating state, these stimuli may gain incentive value to the extent that they will perpetuate behavior for a period of time, and under certain conditions, will act as incentives for the acquisition of new responses. Such stimuli we have called symbolic incentives or rewards.

3. *Primary incentive with symbolic characteristics.* From Brogden's experiment we may make the tentative generalization that a primary incentive will perpetuate a given response for a much greater period of time than will a symbolic incentive, even though that primary incentive does not reduce a primary motivational state. It may be that the visual and olfactory aspects of food in and of themselves are largely symbolic incentives, having gained such reinforcing properties by being consistently present when eating has taken place.

Now we may see somewhat more clearly how it is possible for complex symbolic or derived motives to develop in humans. If we take into consideration the vastly greater symbolic power of humans as compared with animals (this is a human's opinion), it seems quite feasible that our highly symbolic motive-incentive relationships could have developed by processes of learning stemming originally from incentives which satisfied basic physiological motives. We have, however, no experimental evidence other than the type presented here to show how this comes about.

In the chapter thus far we have attempted to set up guides for

thinking about the complex topic of motivation. We have organized this material around the rôles which motivation plays in performance phenomena which are associated with learning. Its application to influence on work will be considered in the special chapter on that subject near the end of the book.

EXPERIMENTAL OPERATIONS INFLUENCING MOTIVATION

We turn now to a consideration of experimental manipulations which are presumed to cause motivation to vary in amount. In each case we shall attempt to discover what effect these changes have on performance phenomena associated with learning. Since many of the experiments have been done with rats a few comments should be made about the composition of groups of rats when each group is to serve under a different experimental condition. In the case of human *Ss*, we have warned against using small randomly selected groups without checking on equality of performance before a variable is introduced. In the case of rats it is customary to form groups without an equality check. Colonies of rats maintained at laboratories become highly inbred so that there is a high similarity in hereditary background. Then, too, the rats have been raised in a common environment. Hence, random selection of groups of rats, even small groups, is likely to result in groups which do not differ significantly. Certain limits are usually imposed on the population of rats from which the groups are drawn. Age range may be restricted in order to reduce the importance of this factor as a source of individual differences. In many instances *E* has made sure that each experimental group contains the same number of male and female animals, and on occasion, he controls the weight of the rats. If these subject variables are actually equalized for various groups the procedure is more similar to Design Method II, Matched Groups, than to Design Method I, Random Groups.

Another method of forming groups of rats is called the *split-litter technique*. A given litter of rats is divided equally among the groups being formed. This method, of course, equalizes age for all groups and assumes a high degree of similarity of hereditary and environmental background. Obviously, it would take more than one litter to fill out the groups in the usual experiment.

Deprivation Time

For our illustrations of the relationship between deprivation time and learning, we shall consider three experiments using rats in a Skinner box.

The gross relationship. Heron and Skinner (148) first trained their rats to operate the bar in the Skinner box. Then the rats were given access to all the food they could eat for a 24-hour period. Following the 24-hour period, they were placed on subnormal rations. Each day the rat was placed in the box and for 4 minutes was allowed to push the lever and get and eat all the food possible. The amount which could be obtained during the 4-minute period was too small, however, to sustain life indefinitely. The procedure of allowing these short periods of reinforcement each day may be called *periodic reinforcement.* After the 4-minute period the rat remained in the box for an additional hour, but during this hour no food pellets dropped as a consequence of bar depression. The number of responses (bar depressions) made during this hour was used as the measure of the influence of the hunger motive. On each successive day up to 5 days, the rate of responding increased. On the fifth day some of the rats pressed the bar over 1000 times during the hour. On the average, the peak of response frequency was reached on the fifth day after which the frequency dropped rather sharply as the rats weakened.

Short term relationship. Perin (285) gave four groups of forty rats each sixteen reinforcements in the Skinner box and then instituted varying deprivation periods for the different groups. A reinforcement is defined as the eating of a single pellet. The degree of learning which took place was shown to be the same for all four groups as a consequence of their having had the same number of reinforcements.

After the training the rats were deprived of food for various intervals to determine how the length of the interval would influence the length of time required to extinguish the lever pressing habit built up by the 16 reinforcements. One group was deprived of food for 1 hour, another for 3 hours, another for 16 hours, and the fourth for 23 hours. Before deprivation started, all rats in all groups were thoroughly satiated. Following the

specified deprivation period the rat was put in the box and left there until no lever pressings had been made for 5 minutes.

The results are shown in Fig. 29. It is assumed, of course, that the longer it takes to extinguish the response the stronger the motivating condition. Thus, the rate of extinction is one of the performance phenomena associated with learning from which we infer different degrees of motivation. The results show that the number of responses made before the rat stopped

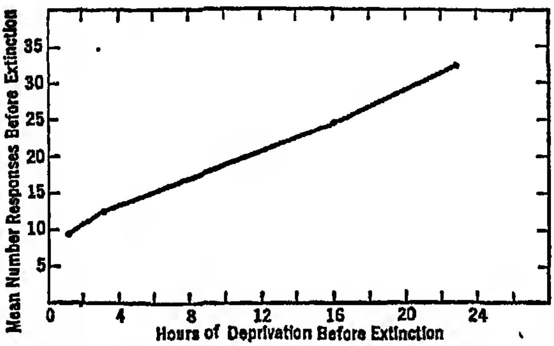


FIG. 29. Relationship between strength of motivation (hr. of deprivation) and extinction time. Each point represents one group of 40 rats. All groups were given 16 reinforcements in the Skinner-box before imposition of the deprivation conditions. Data from Perin (285).

responding for a period of 5 minutes is greater as the hours of deprivation increase. Since the response was originally learned to the same degree in all groups, the differences can be attributed only to differences in the strength of the motive at the time of extinction. In further work, Perin showed that the general relationship holds, regardless of the number of initial reinforcements.

Koch and Daniel (203) extended Perin's work and investigated the resistance to extinction shown when the animals had zero hours deprivation. The rats were given from seventy to ninety reinforcements during training. On the following day the animals were placed in the box immediately following complete satiation. Using Perin's extinction criterion, the results showed that extinction was accomplished in almost zero trials. Actually, the median number of bar depressions made during the

5-minute period was one. It is very likely that we can extrapolate the curve in Fig. 29 at the low end so that it meets the ordinate just slightly above zero.

In summary, these illustrations show that as deprivation time increases, performance associated with learning increases up to the point where physical weakening sets in.

Variation in Incentives

Since incentives are an integral part of the stimulus complex influencing motivation, *E* may vary incentives and determine whether this brings about differences in performance. The process of doing this is one of the most complicated fields in modern psychology, for several reasons:

1. Whether or not incentives will influence performance appears to be dependent upon a great many conditions, and in spite of the fact that there have been scores of researches in this area, there are many divergent experimental results. This is especially true when negative incentives are administered to human *Ss*.

2. Results obtained with animals appear to be different from those obtained with humans.

3. Lively theoretical arguments on the mechanisms of learning stem from this research area because there are several possible interpretations attempting to make plain how incentives might have the influence shown in the data.

For good reasons we are going to postpone all theoretical considerations to a later chapter. We shall not even attempt to review the literature on incentive variation, since adequate surveys are available (26).

Electric shock in human learning. Electric shock for errors is the most widely used negative incentive in experiments involving adult *Ss*. There are many electric shock experiments in the literature which have followed Design Method I, Random Groups. In these experiments all or some of the comparisons between various groups rest on the tacit assumption that the groups were equal in learning ability before experimentation started. No checks were made to determine the validity of the assumption. The reported results from some of these experiments are contradictory. We are left somewhat in doubt as to whether

the contradiction may be due to the fact that two or more groups, assumed to be equal before the introduction of the variable, were not equal.

Let us consider an illustration of how random groups, tested for equality, can be used effectively in studying the influence of electric shock. Gurnee (127) chose five random groups of human *S*s and put them all through four trials on a maze with exactly the same conditions prevailing for all groups. From the scores on these initial four trials he could evaluate the comparability before the variable was introduced. Actually, the mean performance of one of the groups did deviate considerably from the means of the other groups. In the experiment proper he used a control group (no shock); a group given a weak shock for wrong responses; a group given weak shock for right responses; a group given moderately strong shock for wrong responses; and the fifth group given moderately strong shock for right responses. These variations were introduced on the fifth trial and were maintained until *S* mastered the maze. Gurnee found but little difference in the rate of learning among the groups. Within the limits used, intensity of shock did not appear to be a significant variable, and whether it was given for right or wrong responses did not appear to be of any consequence. If motivation was influenced by shock it was not reflected in the performance records.

As a second illustration of sound procedure, let us consider an experiment by Gilbert (111) using Design Method II. Gilbert had 102 *S*s learn a short practice maze. On the basis of their scores he divided his *S*s into two groups so that according to the practice scores they were very well equated as determined by the mean and variability of performance. Then the variable was introduced as *S*s started to learn another maze. One group received no shocks as they learned the maze, whereas the other group was given a shock whenever they had made a certain number of errors —a noninformative shock since it did not occur at the time of the error. Indeed, the shock was often administered between trials. The shock group learned the maze a little faster than the control group, though the differences were not large enough to allow the conclusion that the shock significantly affected the speed of learning.

The procedure used by Gilbert in obtaining equivalent groups makes only one assumption; namely, that performance on the practice maze is highly correlated with performance on the regular maze. This assumption could, of course, be checked by his data, but it is very likely that the correlation would be high. Because the correlation may be somewhat lower than perfect, it is probably better to match *S*s on the actual task to be used by giving the group a few trials before introducing the variable. Practical considerations however often will not allow this.

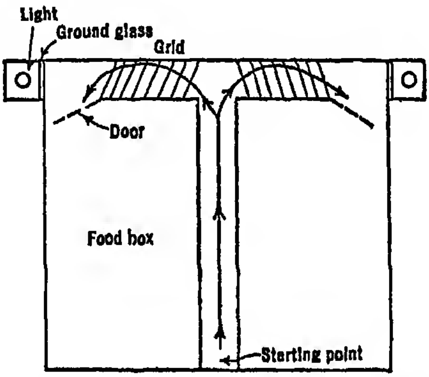


FIG. 30. Plan of discrimination box used by Muenzinger, Bernstone, and Richards (277), in their study of the effect of shock for right and for wrong responses in rat learning. On each trial one of the lights was on and the correct response was going toward the light and entering the food box. On incorrect responses the door to the food box was closed.

We have considered only two experiments concerning the influence of electric shock on learning. They do not, of course, show the wide divergence in results which have been found. There are no general principles which can be drawn from the work done to date except that electric shock usually produces greater variability in performance.

Electric shock in animal learning. With but few exceptions, the introduction of shock in animal learning situations has enhanced performance. One of a long series of experiments by Muenzinger and his students will be used as an illustration (278). In this experiment a visual discrimination box, shaped like a *T*, was used, as shown in Fig. 30. The rat, after reaching the choice

point, learned to take only that alley at the end of which there was a light burning. *E* determined by random order which light should be glowing on successive trials. The animal thus could not learn always to go to the right or to the left; rather, he had to learn to go toward the light. Three groups of rats of twenty-five each were formed. One group received shock from the grid for making *right* responses, another group for making *wrong* responses, and the third group received *no shock*. The rats were given ten trials a day until they made no errors on two successive days (twenty successive errorless trials). The results, as shown in Table 14, indicate clearly that either shock condition was much superior to the non-shock condition in terms of rate of learning. There is some indication that learning was most rapid when the rats were shocked for making wrong responses. The fact that shock for right responses leads to faster learning than does no shock is probably accounted for by the additional discriminative cue (though a painful one) which the shock affords.

TABLE 14

INFLUENCE OF SHOCK FOR RIGHT AND SHOCK FOR WRONG RESPONSES IN LEARNING OF A DISCRIMINATION HABIT BY WHITE RATS

	Errors		*Trials to Reach Criterion*	
	Mean	σ	Mean	σ
Shock right group	17.0	4.1	45.2	16.7
Shock wrong group	11.3	2.7	34.8	11.7
No shock (control)	23.2	6.2	107.2	45.9

Data from Muenzinger, Bernstone, and Richards (278).

Palatability of food incentives. In view of our own preferences, it might seem reasonable to expect that greater energy output would occur if we were working for a very preferred incentive or reward than if we were working for a reward only slightly desired. Children, we might suspect, would perform much better in order to receive ice cream than they would to obtain egg-plant. The restriction of our generalization is rather acute, however, when we attempt to consider adults. In the

human adult we often find the expected relationships obscured by other motives.

If we confine our evidence to animals we find that the relationship between desirability of food and performance rewarded by that food is high. A carefully performed experiment by Young (425) demonstrates the principle. By a cafeteria technique in which animals were allowed to choose between one of two foods (paired comparisons), it was determined that of the three foods used in

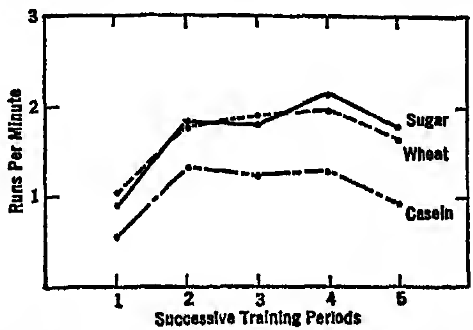


FIG 31. Relationship between performance and palatability of foods. Sugar and wheat were preferred to casein by rats, and are shown here to heighten performance of a simple running task. Data from Young (425).

the experiment, sugar and wheat were almost equally well liked by rats, but both were much better liked than casein. Using these three incentives and a simple task in which the speed of running was the measure of performance, it was shown that the rats made more runs per unit of time when running for wheat or sugar than they did when running for casein. This relationship was determined, of course, when the rats had been subjected to the same number of hours of deprivation. Figure 31 shows the results. The number of runs made per minute is directly related to palatability of the food.

Amount of incentive. The data concerning this dimension are clear-cut for animal experiments. Crespi (73) has shown that the speed of running is directly related to the amount of incentive of a given kind (food). One necessary experimental precaution should be mentioned. In order to vary the amount of incentive among groups, *E* must give no more than 1 trial a day. Since

the amount of incentive is being varied, the degree of hunger will be differentially changed among animals after the consumption of the incentive received on a single trial. Hence, though animals in different groups start with the same strength of the hunger motive the strength will be different for the different groups after the trial. Consequently, only 1 trial should be given a day, and the difference in the amount of incentive received after the trial should be balanced by the amount fed the rats in their home cages. Then for the next day's run degree of hunger would be the same for all groups and differences in performance can be interpreted precisely.

Precise variation in the amount of incentives given humans could be obtained by varying the amount of monetary reward. Again, however, other motives may confound the relationship when adults are used as *S*s. It is not necessarily true that greater motivation will result from the salary of sixty dollars a week than from the salary of fifty dollars. Money would be of no value as an incentive, of course, unless the implications of the money were understood by *S*. Very young children will perform better for candy than for pennies. By and large, as age increases, money takes on incentive characteristics, but because of the complexity of adult motives, we cannot assume that with adults more money is going to bring better performance. These facts are becoming more and more apparent in the case of the relations between labor and management (114, 319).

Praise and reproof. If, following a task or segment of a task, *E* tells *S*: "That was fine," or, "Good, let's do even better this next time," one may think of these statements as being positive verbal incentives, or praise. If *E* says: "Very poor," or, "You're doing a poor job and you ought to be ashamed of yourself," we have negative verbal incentive, or reproof. In view of the fact that many of our "natural" rewards and punishments follow this pattern, it is inevitable that experimental work should be done with such incentives. We need not consider a specific experiment since the method would be straightforward. After equal groups are formed on the basis of preliminary tests, one group would be praised following performance, the other reproved. Usually we would also add a control group which is neither praised nor reproved. Two representative studies have been

performed by Hurlock (174) and Schmidt (335). From these, and other studies, the following conclusions may be drawn:

1. With young children, praise results in better performance than reproof.
2. With older children, reproof seems to be the best incentive.
3. With young adults, neither seems to influence learning a great deal.
4. Either, if given too often, will be reduced in effectiveness, though the loss of the effectiveness of reproof is most rapid.
5. The person administering the reproof or praise is a significant variable —some *S*s may respond to one *E* and not to another.

Summary of incentive variation. 1. Incentives may be varied in kind or in amount of a given kind. Incentives may be considered negative (punishment) or positive (reward). The experimental problem is to determine the relationships between incentive variations and performance phenomena associated with learning.

2. Electric shock has been used a great deal as a negative incentive. For human *S*s the results are ambiguous, shock appearing sometimes to facilitate performance, sometimes to inhibit it, and in still other situations to have no influence. Reasons for the discrepancies are probably numerous, including the type of task used, the degree of dominance of other motives, and the intensity of the punishment. Ambiguity in the data may be due in part to the use of Design Method I. Electric shock administered to animals will usually facilitate performance.

3. Other things being equal, the greater the amount of a given positive incentive the greater the facilitation of performance, and the more desired an incentive the greater the facilitation. These principles are clear-cut in the case of animal studies, but may be masked by other factors in human studies.

4. Praise and reproof, as special forms of positive and negative incentives, have been shown experimentally to influence performance. There are several variables involved.

Competition or Rivalry

"Getting ahead of someone else" appears to be a basic motive in our culture. Very early in life the individual's performance is compared (by himself and others) with the performance of others. Since "getting ahead" is usually rewarded in some fashion,

it is quite conceivable that such a motive could develop. What we call competitive activity is first seen at about the age of three, and then develops very rapidly for the next few years (122). It would appear that in so-called rivalry or competitive situations the incentive is the extrinsic or intrinsic rewards (or both) which usually accompany excelling. Responses to the person and by the person who wins first prize at a bridge party are quite different from those to and by the person who wins the booby prize.

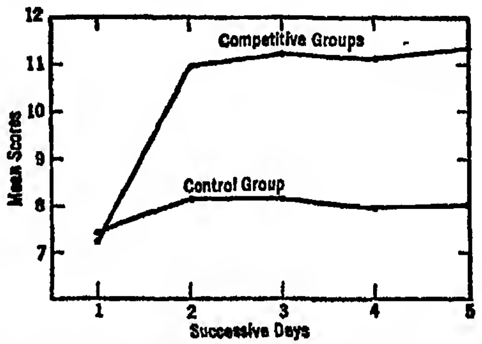


FIG. 32. The effect of competition on the speed of working arithmetic problems among fourth and sixth grade students. The competitive situation was set up on the second day. Data from Hurlock (175).

If we can agree that certain situations involving competition or rivalry invoke changes in the energy level, hence, changes in motivation, *E* may attempt to vary these situations to determine the influence of the variation on performance. We need examine only one study to show that it is possible to influence behavior by changing situations believed to invoke competition. This study was done by Hurlock (175).

Hurlock gave 150 fourth- and sixth-grade students a standardized set of arithmetic problems to solve in a limited period of time. On the basis of the scores, two groups of equal ability were formed. One group served as a control group in which no rivalry was introduced except that which might come about through the usual school situation. The other group, the experimental group, was divided into two sub-groups of equal ability as determined by the initial test. Between these two sub-groups a

competitive situation was set up. To do this, it was necessary only for *E* to announce that she was interested in seeing which group could beat the other group. To add to the stimulation, *E* read the names of the winning *S*s in each group and asked them to stand up.

The experiment extended for four days after the initial division of the *S*s into groups. On each of these days, 30 additional arithmetic problems were given. The control group worked merely as if the problems were another task to be done; the two experimental sub-groups vied with each other for superiority. The results are depicted in Fig. 32 and show that the competitive situation raised the mean scores. Both groups were quite equal at the start (Day 1), but on successive days the experimental groups were consistently and significantly higher than the control.

Induced Muscular Tension

We have insisted that, where possible, *E*'s inferences from changes in motivation should be from changes in the energy output of the organism. The secondary, but very practical problem is that of relating changes in motivation to changes in performance. By definition, we have said that as tension or energy level increases as a result of manipulating a certain class of stimulus variables, motivation increases. We have seen that tension increases during performance of a task as compared with relative rest. Finally, we have seen that performance is related to motivation where it is believed motivation has varied, even though tension measurements were not taken. All of these associated phenomena suggest the possibility that if tension is induced *directly*, performance may be enhanced.

There is a temptation to say that inducing tension increases motivation directly and actually this may be what happens. However, although we may say that changes in motivation are revealed by changes in tension, it does not necessarily follow that changing tension in turn changes motivation. Regardless of how the relationship may be conceived, the problem is wide open to experimental attack and most results to date show that increasing tension will enhance performance within limits. We should consider the methodology of one experiment.

A study by Courts (69) was concerned with the learning of

nonsense syllables (meaningless three letter words, such as *GUP*). Courts, in order to explore varying degrees of tension and their relationship to learning, asked *Ss* to grip a dynamometer as hard as they could for a period of 30 seconds. The reading in kilograms at the end of 30 seconds was taken as the maximum grip measure. Using this maximum grip as 100 per cent tension, Courts required *Ss* to learn lists of syllables while they were gripping the dynamometer at different fractions of the maximum.

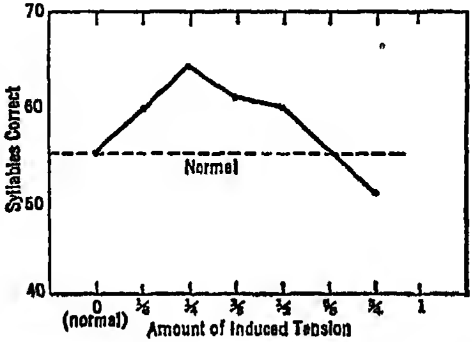


FIG. 33. The effect of induced muscular tension on the learning of nonsense syllables. Data from Courts (69).

The fractions used were $^1/_8$, $^1/_4$, $^3/_8$, $^1/_2$, $^3/_4$. *Ss* also learned under normal conditions, i.e., without gripping the dynamometer. The amount of induced tension was kept constant during learning by placing a small contact switch on the dynamometer dial. When the hand on the dial made contact with this switch a light went on. *S* was to keep the light on at all times while working under induced tension conditions.

Ss had 5 trials on each list of nonsense syllables, learning a separate list for each condition. The results of this experiment are shown in Fig. 33. Note that the optimum point of induced tension seems to be somewhere between $^1/_4$ and $^3/_8$, and that between $^1/_2$ and $^3/_4$ tension, an inhibitory effect begins to occur. In short, inducing tension artificially will enhance performance, as shown by the mean syllables correctly anticipated. With high tension, inhibition of performance may result. It may be noted that some *Es* have suggested that associated emotional factors

may produce the decrement in performance under high induced tension.

The variables in the situation in which muscular tension is induced are not thoroughly worked out. Investigations thus far suggest that the *locus* of the induced tension and the *complexity of the task* are two significant variables (70). A tentative generalization concerning the latter variable is that the more complex the learning task the less the facilitation by induced tension.

The findings with regard to induced muscular tension as exemplified by Courts' study are suggestive of another reason why there have been contradictions in the data concerning the influence of punishment, especially the influence of electric shock. We would suspect that the use of electric shock would increase *S*'s tension. However, some studies have shown a facilitation of performance as a result of shock, whereas others have not. It is possible that the different studies may have invoked different degrees of tension and that this, coupled with differences in the tasks, may have put *S* past the point of optimal tension so that his performance was much as it would have been at a lower tension level. A horizontal line drawn through two portions of the curve in Fig. 33 would indicate points of differing tension level *but of identical performance.*

Variations in Set

Set, a term used in several ways in psychological literature (109), is employed here to relate the *formal instructions* given *S* and performance as a consequence of those instructions. In part, instructions experimentally define the directive function of motivation. They stipulate in what direction *S* should apply his efforts and the stimuli to which he should attend. There is no reason to believe that the instructions as such influence the energizing function of motivation unless those instructions indicate certain rewards or punishments which may follow. Changes in performance due to different instructions, differences which are usually attributed to set, form a part of the study of motivation.

Unfortunately, the problem of set is a complicated one, largely because it is recognized that the instructions given *S* are usually not the only determinants of the direction in which *S* shall apply himself. It has been pointed out (233) that there are at least

three varieties of instructions in the usual human experiment: (1) formal instructions; (2) incidental "instructions" aroused by the situation and without the intent of *E;* and (3) instructions which *S* gives himself. These last two are related directly to *S*'s *past experience*, and how *S* applies himself will be determined to a certain extent by how he has applied himself successfully in the past in similar situations. *E* thus has only partial control over the direction of behavior. Unless set is being studied, of course, *E* attempts to devise experiments so that the past experiences and instructions will not work against each other. When studying set, however, *E* attempts to impose certain conditions on *S* which may or may not be supported by past experience.

"What is learned" as a function of set. In a recent experiment (298) *S*s were given 90 seconds to read a 350 word passage from a dramatic short story. Five different groups were instructed to "read for" different things: (1) general comprehension; (2) specific sequence of events; (3) details of content; (4) details of wording; (5) physical appearance (typographical errors). Three minutes after reading, *S*s were tested over all aspects, i.e., over all five factors, by a multiple choice test.

The results show that some difference in what is remembered can be attributed to the different sets invoked by the instructions, but in no case did a given set preclude learning other factors. Set-inducing instructions did not rule out the acquisition of material not covered by the instructions; hence, this acquisition must be brought about by self-administered instructions, or more likely in this instance, by reading habits of long standing. It is reasonable to presume that one will, to a certain extent, read as he has read in the past and this means reading for "things" for which one has read in the past.

More marked differences can be produced in learning by using less meaningful material so that *S* does not have a response pattern already established concerning the material. In one study (20) a group of *S*s was instructed to *read* a list of nonsense syllables, presented under the guise of a type-clarity test. Another group was instructed to *learn* the list. With a set to learn the list, *S*s took about 5 trials to reach mastery, whereas the group which had been instructed merely to read the list took 12 trials to attain mastery.

Maintenance of set. A variation in the methods for studying set was used by Rethlingshafer (308). She presented *S* with a variety of tasks (nonsense lists, mazes, puzzle problems, and a simple motor task), and while *S* was engaged in learning or performing these tasks the reaction time to a buzzer was measured. *S* was instructed to press a key whenever a buzzer sounded. The onset of the buzzer started a chronograph (timer) and pressing the key (by *S*) stopped the chronograph, thus giving a measure of speed of reacting to the buzzer.

Results show that as the learning of nonsense lists or mazes proceeded, there was an increase in the speed of reaction to the buzzer. The longest reaction times were observed on the first learning trials, and the shortest times were observed when the lists or mazes had been learned and were being performed perfectly. Assuming the puzzle problems allowed for solutions of an insightful nature, the results show that the longest reaction time to the buzzer occurred before insight had taken place, and the shortest after insight had taken place. Finally, during the performance of the simple task (dropping a marble through a hole, picking it up and dropping it through again, and so forth), the reaction time was shown to bc very fast and had but little variability over a considerable period of time.

Note that in this experiment *S* was given two sets, one to react to the buzzer and the other to perform or learn. The results suggest that the more difficulty *S* is having with a given task, the less quickly can he direct his behavior into other channels, i.e., respond to the buzzer. Thus, descriptively it appears that set has "depth," and that this is what we mean by degree of attention.

Control of set by controlling past activity. We indicated previously that *E*'s control of set through instructions was only partial. Set is influenced to a certain extent by past experience in similar situations. We could, then, set up a series of problems whose solutions were similar and thus develop a specific solution set. Next, we could change the type of problem so that then the previous solution set is inadequate, and determine what influence this interference of sets has on the solution of the new kind of problem. Such attacks on the problem of set have been made. The results cut across several areas of study including transfer

of training, interference in learning, and, especially, interference in thinking and reasoning. We shall delay consideration of such set interference until we consider the material on thinking and reasoning.

Ego-involvement

Introduction. In recent years psychological literature has carried an increasing number of references to the term *ego-involvement.* Allport (2), in 1943, summed up current and historical usages of the term and set forth evidence which he believed justified the use of such a term. Recently an entire book, *The Psychology of Ego-Involvements*, has been devoted to the rôle of the ego in its various involvements (351). An extensive discussion of the relationship (or lack of it) between ego-involvement and principles of learning has appeared (310). Inasmuch as some of the problems relating to ego-involvement are amenable to experimental attack, we will need to consider the operations which define it. Our present problem is to examine ways in which *E* can vary motivation. Making *S* ego-involved is presumably one of these ways.

A conception of the ego. We must get some understanding of how the term *ego* is to be used here. Evidence from clinical data tends to suggest the hypothesis that as an individual matures in a given culture, or sub-culture, he gradually builds a rôle (129) for himself in that culture. He develops a status. He learns what is "right" and what is "wrong"; he discovers, with considerable help from his elders, which status or rôle is thought well of and which is considered a failure. He discovers what abilities are highly esteemed, what traits are rewarded, what actions are best to display, and so forth. From a great wealth of experiences he builds up a conception of himself—of his status—which is highly personalized and which includes his feelings of self-esteem and feelings of prestige. This conception which an individual has of his rôle or status we will call the ego. (Cf. Sherif and Cantril, 351, for an extended discussion of the genesis of the ego.)

When an individual is said to be ego-involved experimentally, we mean that he is placed in a situation which is a threat to his status or rôle; it is a threat to his prestige or self-esteem. These threats instigate feelings of anxiety. Anxiety, being a painful

condition, serves as a motivating factor (272, 273) by energizing the organism. His behavior, as a consequence of this anxiety, tends to be directed toward alleviation of the anxiety and thus reduction of the motivating state. The specific direction behavior takes in an effort to remove the anxiety condition will be determined by past experiences which have been successful at this.

With this conception of the ego and its involvement, the experimental problem is that of determining what conditions of threat are universal enough to get a group of *S*s ego-involved by the same operations.

Experimental implications. Allport (2) has pointed out that in the usual laboratory experiment ego-motivated behavior is seldom instigated. He thinks it is a hopeful sign that recent experimentation has attempted to get at these "higher" motives of man. We have already seen that the instructions to *S* may have considerable influence on the performance. It is likely that many of the more staid laboratory experiments *are* unwittingly tapping the ego-motive to a certain extent by *S*'s self-instruction or by false impressions of the experiment. For example, the writer has done some research on the learning and forgetting of simple verbal materials. These experiments are of an individual nature in which only one *S* works at a time. On many occasions, *S*s have asked how highly the ability to learn the words presented is correlated with intelligence, or how highly it is related to personality traits. Without doubt some *S*s have learned the words under the impression that this ability was somehow related to intelligence or personality. Studies of ego-involvement may deliberately use such ruses to attempt to *induce* ego behavior into the experimental situation.

It appears that the recent flurry of writing about ego-involvement has two important implications for the experimentalist. First, it is necessary to reconsider the whole problem of instructions to *S*, since in some experiments we might want him ego-involved and in others not. It is even reasonable to suppose that some of the contradictions in the experimental literature may stem from differences in the amount of ego-involvement unintentionally invoked by the situation and by the instructions. Until more evidence is available, we need not consider the possibility that ego-involved behavior produces qualitatively different

results than non-ego-involved behavior. Allport is careful to point out the dimensional aspects of ego-involvement, and we shall consider motivation changes produced by ego-involvement as extending from low to high in the same fashion that we have conceived the variation in degree of motivation produced by other operations.

Secondly, accepting the concept of ego-involvement poses a new problem in determining what experimental operations can be used to invoke ego-involvement, and what operations can be set up so that it will vary in amount. At present, we may indicate how it is *believed* that one involves egos experimentally. There are two general classes of experimental problems which may be performed.

I. We may give *S* material to learn (or a task to perform), material toward which *S* is not already ego-involved. However, certain conditions are imposed so that the speed of mastery involves ego-motivation. Thus, *S* is given a list of nonsense syllables to learn and is told that the speed of learning these syllables is a measure of intelligence. Few people are ego-involved in nonsense syllables as such, but if they are told that the ability to learn the syllables is a measure of intelligence, they may be highly motivated to learn the list quickly. Failure to do so would indicate to others low intelligence, hence, a threat to self-esteem or prestige.

II. We may give *S* material to learn, the material *as such* being of the nature that *S*'s ego is already involved in it. We would choose one group of *S*s considered to be ego-involved with the material and another group considered not to be ego-involved and determine differences in the rate of mastery. Thus, a group of football players is asked to memorize a passage extolling the virtues of the football team. It is presumed that this group would be more ego-involved in the content of the material than would a group of, say, music-school students.

Experiments using Method II have shown positive results, i.e., the ego-involved group will learn the material faster than the non-ego-involved group. For example, communists will learn pro-communist material faster than will anti-communists (215); pro-New Deal students will learn pro-New Deal material faster than will anti-New Deal students (84). Since, however, such

results may be explained on fairly well established principles of transfer of training (Chapter IX) without invoking any new concepts, we shall say no more about Method II.

Illustrations of Method I. In an experiment by Alper (3), two groups of *S*s were used in the usual control-experimental pattern. Each group had two sessions, but on the second session the experimental group was placed under conditions which were thought to involve a severe threat to *S*'s self-esteem. The main task in each session was that of rearranging jumbled sentences of 20 words. Some of these sentences were easy to place in proper order and some were very difficult, so difficult in fact that they could not be finished within the time limit allowed for each.

During the normal sessions, *E* attempted to maintain a normal, friendly relationship with *S* and presented the tasks as a job to be done. Only *E* and *S* were in the experimental room. During the ego-involved session under which the experimental group served, the atmosphere was considerably different. Several conditions were changed in an effort to pose a threat to *S*'s self-esteem:

1. *S*s were told that the ability to rearrange the sentences comprised an intelligence test used by the Army to select officer candidates. (All *S*s were draft-age males and the experiment was performed during the war.)
2. A male accomplice was situated in the experimental room and he "managed" to solve the problems much faster than *S*, at the same time remarking on the ease of the problems.
3. An attractive female accomplice was situated in the experimental room, ostensibly to help with the experiment, but actually to add to *S*'s embarrassment when he failed to keep up with the male accomplice.

The design of the experiment may be represented as follows:

	Session I	*Session II*
Control Group	Normal	Normal
Experimental Group	Normal	Ego-Involvement

The use of the control group was necessary to show what changes would come about as a consequence of mere practice and as a consequence of the differences in the two sets of sentences, since one set was always used on the first session and the other on the

second. The control group showed no significant change in scores from Session I to Session II, so that any change in performance which occurred for the experimental group may be attributed to differences brought on by the change in conditions at the second session.

The results show that in Session II the experimental group solved *fewer* sentences than they did in Session I. Alper speaks of this as a disruption of performance. Observations made on *S*s, and remarks made by them, suggest strongly that they were highly motivated to do well. Yet, the results show that their performance was poorer under the ego-motivating conditions of Session II than the non-ego-motivating conditions of Session I. The results, then, are quite similar to other results discussed previously; when motivation is very strong, or when tension is very high, performance may be inhibited.

The Alper experiment also has a strong frustration component in it, for, in comparison with the male accomplice, *S actually failed*. The results, then, as far as sheer ego-involvement is concerned, are somewhat obscured. Failure adds an additional factor which we will consider in detail in the next chapter. Ego-involvement as such involves the threat of failure but not actual failure. From Alper's study we cannot answer the question as to whether *S*s would have done better had *E* merely indicated that the performance on the sentence test was a test of intelligence. This is no indictment of the study, however, since Alper was testing other hypotheses.

We must conclude that we have no clear-cut data on the ego-learning of impersonal material as compared with non-ego-learning of the same material. A systematic series of experiments in this area is much needed. There are some data on the relationship between ego-involvement and forgetting, but these data will be considered in the chapter (XIV) on forgetting.

Another class of experiments which falls under this method has come to be known as experiments on the *level of aspiration*. In simple form, a level of aspiration experiment proceeds as follows:

1. *S* is given a task to do, often a simple speed task such as cancelling letters.

2. *S* is asked to estimate what score he thinks he will make

on the next trial on the same task, i.e., what his goal is on the next trial.

3. Another trial is given and the score determined.

The difference between, for example, the score on trial two and the goal set up after trial one reveals something about *S*'s level of aspiration, and may be called a *D-score* (difference or discrepancy score). A series of such sequences may be given and a mean or median D-score determined. If the differences are consistently positive, it indicates that *S* seldom attains the goal he sets for himself; we might say that he over-aspires. If the scores are consistently negative, it indicates that he does better than he says he will do, or under-aspires.

There are many variations on the basic method. It appears, however, that if the goals are set up sincerely by *S*, i.e., if he really expects to make that score, then we have the basic ingredients of an ego-involved situation, since failure to achieve that score would be a blow to his self-esteem. In fact, the logic is sometimes reversed to indicate there can be no true personal failure or success unless the person is ego-involved.

A considerable number of variables in the level of aspiration situation have been worked out (cf. Sherif and Cantril, 351; Rotter, 325; and Lewin, Dembo, Festinger and Sears, 217, for these variables and the results obtained). Strangely enough, there are no data available on the performance of one group working under the level of aspiration conditions and another comparable group working under usual conditions. Consequently we do not know whether the level of aspiration situation increases motivation or not.

Most of the level of aspiration studies are highly similar to studies on competition or rivalry, either competition within the group or of an individual with himself. Most of the interest to date has been centered on changes in performance and on changes in aspiration scores which occur within a series of trials. The above references should be consulted for these facts. It may be that D-scores in a variety of tasks will be found to have some significance as indicators of degree of adjustment. D-scores show a wide range among individuals along a reality-irreality dimension (179). It would appear that a person who consistently sets his goals higher or lower than his performance would objectively

warrant (irreality) has a different personality structure than does a person whose aspiration scores keep pace with his performance (reality).

Studies of the level of aspiration clearly fit into the study of motivation since they are concerned with goals and with attempts to attain these goals. Indeed, in these studies we have the rare case of laboratory behavior in which *S* chooses his goal and then tries to attain it. This choosing, and striving to attain, fits our conception of ego-involvement, but we must await considerably more research before the data obtained by the level-of-aspiration technique can be integrated with the rest of the topics of motivation.

Summary. All in all, in this section, we find that for the experimentalist ego-involvement remains fairly resistant to laboratory manipulations. We must admit, then, that our derived motives, mentioned earlier in the chapter, still remain quite obscure as far as defining operations by which we can manipulate them are concerned. The operations discussed here are the only ones available to date. Undoubtedly future research will result in more definitive data on the relationship between ego-involvement and performance. Of this much we may be sure: there are enough psychologists who are ego-involved in the concept of ego-involvement so that further research will be undertaken.

SUMMARY

1. Motivation is an hypothetical process which we infer from certain changes in responses occurring as a result of certain changes in the stimulus conditions. The response dimensions from which we infer changes in motivation are two in number:

a. changes in tension or activity level
b. changes in performance

2. Learning and motivation are closely related. Certain problems relating to motivation were discussed under the three functions which motivation is assumed to play in the learning process.

These three functions are:

a. energizing the organism
b. directing behavior
c. emphasizing or selecting responses

3. The problem of motivation as investigated in the laboratory as well as observed in everyday life includes the goal objects or incentives toward which behavior is directed. Other stimuli occuring consistently with a primary incentive, such as food, may also gain power to direct or sustain behavior for limited periods of time. These are called *symbolic incentives.*

4. We have certain basic primary motives and others which we have called *derived motives.* In working with animals we employ primary motives; with human *Ss* we are working largely with derived motives. The derived motives are presumed to be built upon primary motives by a process of learning similar to, but much more complicated, than that from which symbolic incentives are developed.

5. Classes of stimulus manipulations which are believed to influence motivation were discussed. These manipulations are:

a. deprivation time (used with animals)
b. incentive variation
  (1) negative incentives (punishment)
  (2) positive incentives, including kind, amount, and desirability.
c. competition or rivalry
d. induced tension
e. variations in set by instructions
f. ego-involvement

6. Three areas, especially, are in need of much more experimental work:

a. negative incentive variation in human *Ss*
b. induced muscular tension
c. ego-involvement

CHAPTER VII

Frustration

INTRODUCTION

Frustration is common in our society. It is revealed in the behavior which accompanies attempts to get something that is difficult or impossible to get. The social climber, uninvited to parties she would like to attend, indulges in malicious gossip; the golfer who shoots 90, and who thinks he is able to shoot 80, throws his clubs in the lake; the student who tries to get *A*'s, and who consistently gets *C*'s, may change his major and forever condemn individuals in the area in which he had shown incompetence.

Definitional problems. The study of frustration is an elaboration of the study of motivation, since it is concerned with behavior resulting from failure to satisfy a motive. Organisms are motivated in certain directions, i.e., motivated to solve a problem or to reach an incentive or goal-object. If the problem situation is an uncomplicated one, the organism, when aroused, "unfolds" a behavior sequence previously acquired and attains the goal-object. If the organism is in a relatively new situation for which it has no appropriate response immediately available, trial behavior occurs and the organism usually chances upon an adequate response. If in either of these situations (during the execution of a learned response or during the acquisition of a new one) blocking or interference with goal-directed behavior occurs, we have the condition which may produce frustration.

To define the *stimulus changes* necessary to produce frustration we refer to *blocking* or *interference* with a goal-directed behavior sequence. This is not an operational definition, however, for if we ask when an organism is frustrated, or ask what are the criteria used to determine when frustration is present, we run into danger

of circularity. Suppose, for example, we block or interfere with goal-directed behavior and observe responses which are not different from those which occur in the absence of blocking. Are we to say that this is not a frustrating situation? If we say this we lose the meaning of our definition, for now we would be denying that the blocking of a goal-response is frustrating. To avoid this misconception we must again revert to an examination of the experimental operations used to study the phenomenon.

To produce frustration E blocks or interferes with goal-directed behavior. It is observed that, after blocking, the behavior of the organism deviates from or is not the same as the behavior which occurred when no blocking took place. This deviant behavior, *deviant compared with that in a non-blocking situation,* is what is called *frustrated* behavior. To define frustration in terms of the operations by which it is measured, we have to correlate both the stimulus manipulations and the response measurements. Consequently, we must define frustration as *that which leads to the deviant behavior which is observed as a result of the blocking of or interference with a goal-directed behavior sequence.* Thus, frustration is another logical construct. When E blocks goal-directed behavior, and when such blocking results in deviant response (compared with a control group not blocked) we infer a state which is called *frustration.*

If we put the definition in terms of an experimental and control group, it would appear as follows:

Control Group: No Blocking
Experimental Group: Blocking

The behavior of the control group provides basal measurements, and any deviation from these measurements shown by the experimental group must be attributed to the difference in stimulus conditions, i.e., the blocking. It is unlikely that we have many conditions in which there is *no* blocking, so that difference between a control and an experimental group is one of difference in amount of blocking. If E believes he understands the behavior which occurs when blocking is relatively small in amount, he may not use a control group. The control group is always implied, however, in making inferences concerning the influence of the blocking.

Background of research. Experimental work concerned with frustration and conflict (next chapter) is recent in origin, the bulk of it having been done during the last 15 years. Probably no other area has so much intrinsic interest for as many psychologists as does this area. For several reasons the work has special significance to psychologist and layman alike.

1. An experimental analysis of the conditions producing abnormal behavior may shed light on the causal mechanisms producing the behavior with which the clinical psychologist and psychiatrist must continually deal.

The work done thus far is at best only suggestive, and it will be well to caution ourselves on two points: (1) some of the work has been done on animals and, as already indicated, we cannot generalize to human behavior without reservations; (2) when minor behavior disturbances are produced experimentally in human *S*s we cannot be absolutely sure that this behavior is on the same continuum with abnormal behavior as defined by the psychiatrist. All we can say with regard to both points is that research results thus far seem very encouraging and there is no reason to quit work because reservations concerning its interpretation must be entertained. Whether we are getting at the same mechanisms as those present in serious abnormalities or not, the data collected embody important laws of behavior; thus, experiments in the area will be continued.

2. Freud's writings have had tremendous influence on thinking about human behavior. Since his work first became generally known to American psychologists there have been conflicting feelings concerning it. It was as though psychologists intuitively felt that many of Freud's insights, incorporated in such clinical concepts as repression, fixation, and so forth, were sound. Yet, for lack of what they would call reliable evidence, they were not gullibly accepted by objective-minded psychologists. Later experimental work has helped resolve some of these conflicts by showing that some, at least, of Freud's clinically derived mechanisms or concepts can be empirically substantiated. (See Sears, 338, for a comprehensive survey of objective data as related to psychoanalytic concepts.)

3. Some experiments in this area have been concerned with group responses. That is, the responses observed were those of

a group of *S*s as a whole. These observations, plus those on individuals, have provoked rather extensive theories to account for a wide range of group behavior, such as wars, lynchings, fascism. The theories have been experimentally oriented, and rather ingenious methods have been devised for testing the theories.*

Terminology. We have defined *frustration* as a state we infer when deviant behavior follows the blocking of goal-directed behavior. In the previous chapter we defined *motivation* in terms of energy output. We must now state the manner in which the concept *emotion* is to be used, and how it is to be related to other concepts.

Usually descriptions of frustrated behavior include changes in emotional tone. The terms applied descriptively to emotional states fall into two classes: (1) names indicating increased emotional tone, e.g., fear, anger, hate, excitement; (2) names indicating decreased emotional tone, e.g., apathy, dejection, ennui, lethargy. *For our purposes these reported changes in emotional tone may be considered changes in motivation.* In our discussion the terms may be used interchangeably, with high emotional tone names meaning high motivation, and low emotional tone names indicating low motivation. This conception may run counter to well-established habits of thought. At present, all that is asked is that the above point of view be kept in mind as we examine results from frustration and conflict experiments. At the end of the next chapter, after we have had opportunity to examine experimental evidence, we shall detail reasons for the identification of motivation with emotion.

Plan of the chapter. One of the major methodological problems in studying frustration is analyzing the responses to determine whether or not behavior deviations are produced. The bulk of the present chapter will be concerned with this problem. Initially, however, we must get an over-all picture of methods which have been used to block goal-directed responses, and we must also preview the responses for which *E* usually looks. The chapter follows this sequence:

* Most of this experimentation has stemmed from two sources, one centered at Yale University, and the other around the late Professor Kurt Lewin and his students. The work with groups of *S*s has commonly been called *group dynamics*.

1. Methods used to produce frustration
2. Major responses to frustration
3. Methods of response measurement and analysis
4. Stimulus variables influencing responses to frustration
5. Final considerations

METHODS USED TO PRODUCE FRUSTRATION

Interference with primary motive. Anyone who has tried to take a bone from a hungry dog knows that even the friendliest animal may show considerable irritation at this treatment. This behavior results from interference with goal-directed behavior. The most nearly analogous experimental procedure with humans has consisted in literally taking milk from a baby. Hungry infants were allowed various amounts of milk and then the bottle was taken away (248). The infants' reactions were observed to determine the relationship between strength of motivation (assuming that the more the baby had drunk the less the motivation) and reactions to the interference.

In another experiment interference with a primary motive was accomplished by forcing adult *S*s to stay awake all night under rather trying circumstances and then noting their various reactions (341). The sleep motive can become very powerful, and interference with its satisfaction may result in deviant responses indicative of frustration.

Physical obstructions. This method is not easily employed with adult *S*s, but is commonly used with children and animals. Even so, one experiment (101) has been reported in which adult *S*s were frustrated by physical barriers. These *S*s (in groups) were placed in an attic room. Unknown to them the exit door was locked. *E* arranged to have a fire "break out" and then observed reactions through a one-way screen in the roof. This situation is probably the nearest to a panic situation that has ever been set up in the laboratory.

In working with children, *E* places a physical barrier between the child and a much-desired object. In one study (10) children were allowed to play with a collection of new and interesting toys for a few minutes. Just after the child had become engrossed in the new toys, he was removed to the other end of the room and a wire-mesh screen dropped as a barrier preventing his

reaching the toys, but not from seeing them. In another instance the toys were placed under a box which was too heavy for the child to move (188). Most children will not remain calm and dispassionate in such situations.

In working with animals the obstruction technique is common. For example, young chimpanzees were taught to find food in cans moving along an endless belt. After this "expectancy" was well established a trap door would close over the can just as the animal reached for the food (143).

Note that in the above situations *Ss* are totally prevented from reaching the goal. This represents a major difference between a frustration situation and a simple learning experiment. In a learning situation, such as a maze, there are obstructions between the animal and the food, but by appropriate search the animal can eventually reach the food. Thus, the usual learning situation may be thought of as one in which the organism is temporarily blocked and the frustrating situation one in which duration of blocking is somewhat longer.

Artificially producing failure. With human *Ss*, especially adult *Ss*, producing failure usually involves, interference with the self-esteem motive, or, if you prefer, the blocking of an ego-motive. There are several ways by which this may be accomplished.

1. *Setting Unobtainable Standards.* This method uses situations in which it appears to *S* that he should be able to complete a task or solve a problem, but actually he can't. In one experiment a tapping task was used (1). In this test *S* held a metal stylus in his hand and attempted to tap a steel plate as rapidly as possible (the movements were vertical). Each tap completed a circuit to activate a counter. To produce frustration *E* set another counter so that it would revolve at a speed slightly greater than *S* could drive his counter by tapping at high speed. *S* was instructed to keep up with the pacer counter.

In another experiment children were instructed to learn a series of numbers (186). The numbers were projected on a screen, one at a time, with *S* trying to anticipate the number before it appeared. The series was repeated until *S* could anticipate correctly all the numbers in the series. This training program continued with the number series gradually being made more difficult by increasing their length. *S* was able to learn each series, however.

Finally, on the frustration day, *E* made the series still longer and then from trial to trial changed certain digits in the middle of the series so that it became impossible to learn. Since the series was quite long, and since *S* had been successful in learning on previous days, *E* was able to do this in most instances without *S*'s suspecting that he was being hoodwinked.

2. *Falsifying Scores.* This method may be used in conjunction with the level-of-aspiration technique (see previous chapter). *S* is asked to indicate what score he is going to make on the next trial. *E* then reports to *S* that he (*S*) failed to score as high as his stated aspiration. For example, in one experiment *S*s sorted playing cards into four boxes, one for each suit, as fast as possible. On each successive trial they were asked to estimate what score they hoped to make on the succeeding trial. Regardless of the score actually made by *S*, *E* consistently reported it lower than the aspiration score (336). The method may also be modified by telling *S* that he must make a certain score before the experiment is complete, and then by falsely reporting the scores, seeing to it that *S* never reaches that score.

3. *Condemning S as Inferior.* When *S* is made to fail by falsifying scores, frustration can be enhanced by *E*'s condemnation of *S* for doing so poorly. Thus, in an experiment on card sorting, *S* was forced to fail by falsifying scores, and after each trial *E* would berate *S* for his stupidity and for spoiling his experiment (227). This method presumably can be very effective if adequate self-defensive measures are taken by *E*.

4. *Combinations of the Above.* *E*s have compounded the frustrating situation by using several methods simultaneously. In one procedure (412) *S* was given a relatively simple task to perform under the following conditions:

a. *S* was told that the task was a measure of intelligence. (When *S* was caused to fail it would amount to calling him stupid.)

b. *S* knew he was being observed through a one-way screen by a number of psychologists who were ostensibly making notes on his behavior.

c. *S* was sitting so that photo-flood lights fell directly on him; he was told that all reactions were being photographed.

d. *S* was told that he would be given electric shocks whenever his performance fell below "our standards." Then he was given shocks periodically regardless of how well he did.

Procedures (*b*) and (*c*) above require that *S* perform in a situation which is extremely distracting. *Extreme* distraction may by itself be used to produce frustration even without such additional factors as telling *S* that the task measures intelligence, that people are watching, and so forth. The requirements are only that *S* be motivated to reach a goal and then encounters conditions which markedly retard or prohibit him from reaching that goal.

Creating a conflict situation. When stimuli for two incompatible response tendencies are present simultaneously, we have a conflict situation. We will consider conflict one means of inducing frustration, and because of the distinctive nature of the work we will allot the entire next chapter to its study.

These, then, are the major methods which define the stimulus situations evoking deviant responses and allow us to say that frustration is apparent. Details of the methods will be brought out in the material to follow.

MAJOR RESPONSES TO FRUSTRATION

There is no general agreement on what behavior should be expected in a frustrating situation. No classification of responses has been generally accepted. However, there are certain mechanisms of adjustment (learned response patterns) which arise in most frustrating situations and with which we should be familiar before examining the methods of measurement. Knowing these, we will at least have an idea of what *E*s are looking for. Some of the names have their origin in psychoanalytic literature, and the experiments reflect in part the pursuit by experimental psychologists of objective evidence for the mechanisms derived from clinical case studies. We will indicate only the major mechanisms and give a brief description of the behavior they symbolize.

1. **Aggression.** In its most restricted meaning this term refers to the tendency for *S* to attack the source of blocking. It indicates hostility and is often accompanied by emotional states of anger or hate. We will, however, use the term in a broad sense. In our society aggression will often take much more subtle forms than those of direct attack. The aggression may be displaced or projected away from the actual source of blocking.

In a theoretical outline to guide research in the area of frustration,

a group of Yale University psychologists postulated that inhibition of direct aggression is due to fear of punishment (81). While it has been difficult to get experimental verification of this principle, there is every reason to believe that it represents a fundamental relationship. Anecdotal evidence will give abundant support to the hypothesis. Enlisted men seldom strike an officer, yet as many can testify, there have been situations in which the impulse to physical attack was certainly powerful. For some reason this impulse was inhibited, probably by the knowledge or experience of the punishment which one receives for such attacks. Overt physical aggression, of course, will seldom break out in the experimental situation. Instead, the aggressive impulses will show themselves in indirect ways.

2. **Withdrawal.** Another way of reacting to a frustrating situation is to get away from it. This may involve actual physical withdrawal or it may take the form of a psychological withdrawal. Of the latter, we may identify three kinds.

a. Regression. Regression means a "going back". This refers to the fact that organisms under stress may revert to a behavior that was appropriate at an earlier age level. Extreme cases of psychotic regression show adult patients behaving as children in their play habits and speech. Sheer physical withdrawal itself may be a form of regression. Needless to say, such extreme behavior as found among psychotic patients is not produced in the laboratory.

A laboratory definition of regression allows one to point out specifically the behavior to which the organism reverts. This is accomplished by requiring the organism to learn one habit as a response to a situation, then replacing it with another habit. Following the acquisition of the second habit the organism is then frustrated in (prevented from) making this second response. Regression takes place if the first learned habit reinstates itself.

b. Apathy. This word suggests that motivation to obtain a goal is reduced. It is a form of motivational deterioration and will occur after rather prolonged failure. Excessive day-dreaming is often found as a concomitant response.

c. Repression. This word is used to describe the inordinate forgetting which may result from anxiety-producing situations. It is as though painful anxieties are avoided by "ignoring" the situations which evoke them.

3. **Fixation.** In certain frustrating situations, usually involving intense punishment, response tendencies may become very strong and rigid. Since responses do not become as resistant to change in the ordinary learning situation, these fixations may be considered deviant responses.

4. **Miscellaneous.** The above mechanisms are the common ones produced experimentally. A few miscellaneous mechanisms will come up in our discussion. Most, but not all, mechanisms are adjustive in some degree because they tend to protect the individual from failure in one way, if not in another. *Rationalization*, for example, tends to accomplish this by providing excuses for behavior. On the other hand, blocking of goal-directed behavior may in some cases result in excited, tense behavior which shows little finesse or rationality. Problems are attacked without deliberation as if in effort to solve them by sheer force. Such a response may be described as a *bulldozing response* and appears to be relatively non-adjustive.

It must be clear that the above listing of response mechanisms is not intended as a classification system. The concepts are offered only as a descriptive guide to help us in understanding the results of experiments which will be presented in this and the following chapter. Operational definition of the mechanisms will be given via description of the methods used to measure them.

METHODS OF RESPONSE MEASUREMENT AND ANALYSIS

Unfortunately, the responses which result from frustrating situations are not simple by any means. They are not always behavior segments which can be categorized and therefore counted; they are seldom right or wrong, nor can they be recorded by mechanical means. Experimentation in this area is still largely in the exploratory or I-wonder-what-would-happen stage. Several techniques of measurement have been tried; some have proven successful, others not. This section will give an over-all perspective of these methods.

Analysis of Verbal Responses

An experiment by McClelland and Apicella (227) illustrates a method of analyzing responses to frustration entirely on the basis

of verbal behavior. The purpose of the experiment was to produce and classify as wide a range of verbal responses to frustration as possible. First let us understand the stimulus situation.

The immediate *E* in this experiment was an undergraduate student who stood in somewhat prestigeful position with other students. He asked a group of them to serve in his experiment, to help him get through a course in psychology. Some were good friends, others were only acquaintances of *E*. In any event, "the subjects were motivated to perform in such a way as to please him" (p. 377).

The task used was card-sorting. *S*s were told that they would finish whenever they were able to sort the deck in 45 seconds. The level-of-aspiration technique was employed, *S* estimating what score he would make on each successive trial. *E*, in reporting the scores to *S*, falsified them so that on two out of every three trials *S* did not attain the aspiration score. This alone is supposed to bring feelings of failure and frustration. In addition to this, however, *E* made sarcastic comments about *S*'s poor ability. These comments were of four types and are worth quoting to get a clear picture of an unusual experimental method. The remarks refer to what *E* said to *S* at various times during the card-sorting trials.

a. Calling attention to failure, e.g., "Not too good, you screwed that one up, failed that time, that's one failure." These comments grew more intense as the failure continued. "That's four times you flunked. That stinks."

b. Comparing the subject, to his disadvantage, with the group, e.g., "You're far the worst I've had yet."

c. Complaining that the subject is ruining the experiment, e.g., "You're ruining my whole experiment, G—— d———it! G——, I'll never pass this course with these lousy results."

d. Becoming aggressive toward the subject and attacking him directly: "You dumb b——. You failed, you stupid s——." (p. 378).

Two groups of 14 *S*s each were run. The second group was subjected to even more severe comments than those indicated above. This was done in an attempt to produce a greater variety of responses.

The responses recorded were the verbal reactions of *S* during and between the card-sorting trials. This recording was done by

E in the experimental room and by another *E* who listened through a speaker attached to a hidden microphone in the experimental room. Of the 28 *Ss*, all but one gave verbal reactions which could be classified. The mean number per *S* was 6.46.

After examining the responses, *E* devised a classification system for categorizing the recorded verbal reactions. The implication of each statement was studied and placed in one of the various categories. The question immediately arises as to the reliability of the classification. Unless there is high agreement among different judges, the implication of the response is statistically ambiguous. Although no thorough check was made by having several neutral judges classify the responses, the two *Es* agreed very closely concerning the categories into which reactions were placed.

We shall not reproduce the entire classification table and illustrations as they are given in the original report, but a small portion is shown in Table 15.

TABLE 15

ILLUSTRATION OF BEHAVIORAL CATEGORIES AND REMARKS SO CLASSIFIED IN THE MCCLELLAND-APICELLA EXPERIMENT (227)

Aggressive Responses	*Sample Remarks*
Directed at Self	"G——, I'm good!"
Directed at *E*	"I know I failed you s——!"
Directed at Cards	"Come here, you b——!"
Directed at Psychology	"Thank G——, I don't take this stuff!"
Withdrawal Responses	"That's about the best I can do." "I can't get it." "Have I got to stay here 'til I get 45?" "You couldn't take me away from this for awhile?"
Rationalization	"I could do better if I couldn't see you." "I didn't get enough sleep last night." "It's too hot."

According to our basic definition of frustration we must include some norm from which deviations are measured. This would be a control group working at the same task but having no frustration techniques applied to them. A third group of 14 *Ss* served as such a control in this experiment. While the classification of the verbal reactions of this group is not given, it is mentioned

that only 26 verbal reactions were given by all *S*s in the group. Compared with the mildly frustrated group which gave 50 responses, and the strongly frustrated group which gave 131 reactions, there can be no doubt that the technique produced deviations in verbal behavior. It is presumed that not only is there a deviation in number but also in kind and intensity.

The frequency with which a given kind of reaction is reported depends upon the reliability of a classification system as well as its structure. Because of this, amounts and kinds of reactions to frustration in any given experiment must be interpreted cautiously; also, inter-experimental comparisons are almost impossible. In this particular experiment, by combining categories to allow a liberal definition of aggression, it was shown that 40 per cent of the responses in the first group were aggressive in nature as were 48 per cent in the second group—the one most severely treated. About 12 per cent of the responses were classified as withdrawal responses in both groups, with the remainder scattered among several other categories.

Analysis of Verbal and Motor Responses Combined

In the previous experiment behavior was analyzed on the basis of verbal responses, with little attention paid to other behavior. Since a verbal response may have different meanings in different contexts, additional aid in interpreting the reactions should be given if non-verbal behavior is recorded simultaneously. This is the method used by Zander (426) on 34 fifth- and sixth-grade children.

The method used by Zander consisted in presenting *S* with a problem which he would fail to solve because the problem was insolvable. For three days *S* learned series of digits flashed on a screen. One digit at a time was shown and *S* was given 3 seconds to guess what the next number would be. By repeating the series, *S* could anticipate the numbers before they were shown. On each of the three practice days which consisted of 15-minute work, *E* made the situation such that all *S*s were successful, i.e., they learned all series. The length of the series was increased gradually on the first three days until on the fourth day it was made long enough so that *E* could change a digit or two in the middle of the series from trial to trial without *S*'s suspecting. This, of course,

forced failure. If we consider the first three days as control days, and the fourth day as experimental day, comparison of behavior on the control days with the experimental day should give a measure of frustration.

The behavior of *S* was recorded by an observer seated in the experimental room. The observer was known to the children, and since most of them had been used in other experiments, they were accustomed to being watched and tested. The observer wrote down everything *S* did and said. In addition, three "time samplings" of gestures and other signs of bodily stress were made during each experimental period. Each sample was 1 minute in length, taken during the 2nd, 8th, and 14th minutes of the 15-minute period. Time sampling is based on the same theory upon which all sampling is based; namely, truly random samples will allow one to generalize to the population (total behavior) from which the sample is drawn. Both the general observations, taken throughout the 15-minute period, and the time-sampling data were used in evaluating the effect of the stimulus situation.

In analyzing effects of failure, the protocols for each day for each child were scrutinized, with verbal behavior taken in conjunction with non-verbal behavior. A classification permitting the calculation of per cent of *S*s showing a given kind of behavior was evolved. *E* made the judgments concerning classification. Some of the data are shown in Table 16.

TABLE 16

PER CENT OF *S*s SHOWING VARIOUS REACTIONS TO FRUSTRATION IN ZANDER'S EXPERIMENT (426)

Behavior	*1st Day*	*2nd Day*	*3rd Day*	*4th Day*
Aggression	20.0	21.2	25.5	25.9
Withdrawing	5.7	5.7	10.1	34.8
Non-Adjustive	38.4	34.2	36.0	44.2

Withdrawing behavior was assumed in cases where there was inattention, pouting, crying, and reduced volume of voice. Non-adjustive behavior which is much the same as that which we have called bulldozing behavior was illustrated by blushing, sighing, intense effort, and general bodily tension.

It will be noted in Table 16 that there is marked increase in withdrawing on the fourth day and some increase in non-adjustive behavior. Zander indicates that the coöperation with *E* dropped a great deal on the last day, indicating an indirect manifestation of withdrawing behavior. The situation took on some frustrating aspects even before the fourth day. For the most part the children did not like the task and actively resisted serving as *S*s. Table 16 shows little evidence of an increase in aggression on the fourth day. However, Zander presents information suggesting that aggressive tendencies were present even though they were not evinced in the experimental room. For example, one boy, after the fourth session, returned to the classroom and immediately started a fight, although he had shown no aggressive behavior in the experimental room and the teacher indicated that starting fights was alien to his usual behavior. It is worth noting also that Zander became extremely unpopular among the students in the school. To quote him:

> Individual cases of hitting the writer, making faces at him, calling names, aggressive remarks, and even kicking him were frequent during the weeks of the study, happening at least once a day as he passed through the class room or the hall (p. 26).

It is difficult to establish definitely that the source of these relatively uninhibited aggressions was the frustrating experimental situation to which the children had been exposed, but it is a reasonable hypothesis. One class became so unmanageable while serving in the experiment that it was deemed wise to use no more *S*s from the class. Thus, while the records taken in the experimental room do not show any progressive changes in aggression, nor any sudden changes on the fourth day, there is reason to believe that aggressions were present which "worked themselves out" outside of the experimental room.

Measurement by Projective Tests

We have discussed the reason why aggressions may not be openly displayed in our society. Projection is a mechanism which allows for the displacement of aggression; it is presumed to allow *S* to "disown" the tendencies. For example, a boy says that a neighbor's son is trying to "beat up on him." This might indicate

the attribution to another of aggressive tendencies which he himself feels. Descriptively, it is as if by allotting the aggression to others, he himself gets rid of it. In a previous chapter (V) we have seen how motivation may influence one's perception. What one sees is not solely a function of the objective situation. Now, if there is a tendency to displace aggression because direct, overt aggression is punished in our society, and if frustration produces aggression, and finally, if our perception is influenced by such aggressive tendencies, it might be worth while to frustrate individuals and provide a standardized medium through which aggressive tendencies might be filtered. The projective tests may be used for such purposes.

The best known of projective tests is the Rorschach Ink Blot Test (14). *Ss* are shown a series of ink blots and are asked to indicate what they "see" in these blots. After such a test is given to a great many people, norms can be prepared, and responses which are rarely given may be considered deviant responses. Another projective test is called the Thematic Apperception Test (266). This test consists of a series of pictures, usually suggesting a traumatic situation. For example, one picture shows a surgical operation in progress, with a young man the dominant figure. Another shows a lad on a couch with a pistol nearby. *S* is asked to write or tell a short story around each picture. Whether or not the story-response is considered deviant depends upon how much the central theme of the story differs from stories told by others. The Thematic Apperception Test was used in the following experiment conducted by Bellak (15).

Ten pictures and 10 adult *Ss* were used. *S* was given 5 cards, failure was introduced, and then 5 more cards were given. Will the themes of the first 5 stories differ from those of the second 5? In order to get a scientific answer to this question, we would have to consider the story-evoking powers of the cards as such. Do we know that the five cards first used evoked stories comparable to those evoked by the second five? Unless this problem is considered, differences in the stories before and after frustration might be ascribable to the cards as such and not to the frustration. There are three ways to handle the problem: (1) Two equivalent sets of cards may be discovered by empirical try-out. If it can be shown that themes written around the two sets are fundamentally

equivalent, one set can be given before frustration, the other after. (2) If it is shown that the same set of cards is responded to in approximately the same fashion on two successive occasions without intervening frustration, or if the amount of change from one testing to a second testing is known, then the same set of cards could be used before and after frustration. (3) The cards may be counterbalanced, as was done by Bellak. Five *S*s were given cards 1-5 before frustration and cards 6-10 after frustration. Five other *S*s were given cards 6-10 before frustration and cards 1-5 after frustration. Thus, each card was used equally often before and after failure, and differences in the responses for all 10 *S*s combined cannot be attributed to differences in the cards themselves.

After the start of the experiment, the following instructions were given *S*:

> This is an opportunity for free imagination. I want you to make up a story. Tell me what has led up to the situation shown in the picture, describe what the characters are feeling and thinking, and tell me what the outcome will be. Speak your thoughts out loud as they come to your mind. Use your imagination freely, and make up anything you please (p. 356).

After each *S* had told stories for 5 cards, *E* frustrated, or at least attempted to frustrate him by telling him that the stories were almost the worst he had ever heard. Then the other 5 cards were given.

E's measure of deviant response was the number of aggressive words given in verbatim recordings of the stories for the last 5 cards as compared with the first 5. These stories were given to two judges who did not know about the experiment, with instructions to count the number of aggressive words in each story. The mean number of aggressive words before frustration was 12.6, after frustration, 23.7. Testing this for significance a difference so large would be expected about once in 20 times if the true difference were zero. It would appear that the frustrating situation, imposed after the fifth card, had produced a reliable increase in the number of aggressive tendencies as measured by the expression of these tendencies throughout the stories. Since somewhat similar results have been found by Rodnick and Klebanoff (318), we may consider the projective-type test a promising instrument for measuring reactions to frustration.

Measurement by Questionnaire and Ratings

This method has been used primarily in the hope that it would measure displaced aggressive responses. The rationale of the method seems sound. Supposing, for example, *S* rates a series of professions according to the esteem in which he holds them. Let us indulge in a little wishful thinking and say that psychology ranks at the head of the list. Now, this group of *S*s is subjected to a very frustrating situation by a psychologist. Immediately after, *S*s are asked to rank the professions again. It is reasonable to suspect that psychology would no longer be at the top. The assumption would be that frustration by a psychologist generated aggression toward psychologists as a whole and that this would be reflected in the ratings.

The method, however, has not always been successful. We shall consider one illustration in which the method has proven discriminating and another in which it has not.

Displaced aggression in an attitude questionnaire. The 31 *S*s used by Miller and Bugelski (261) were workers at a summer camp, their ages ranging from 18 to 20. The frustrating situation to which they were subjected is described as follows:

> As a part of the educational program of the camp they were going to be required to take some long, uninteresting tests most of which were so difficult that everyone would be bound to fail miserably. Furthermore, the tests would be certain to run far overtime so that the young men would miss what they looked forward to as the most interesting event of the otherwise dull week: bank night at the local theatre, which was awaited with special eagerness on the occasion because one member of the group had won $200.00 the preceding week (pp. 437-438).

An attempt was made to measure the displaced aggression generated in this situation by determining the attitude of *S*s toward Japanese and Mexicans before frustration and again after frustration. Even before the war, when the experiment was performed, Mexicans and Japanese were considered by many to be out-groups and prejudice existed toward them.

Attitudes toward these two groups were measured by a check list of 20 traits, 10 of which were desirable and 10 undesirable. *S*s were asked to indicate a plus if they considered that these

traits were present in the average Mexican (or Japanese). They were not to check the words which didn't apply. Examples of the desirable traits are *friendly*, *smart*, and *patient*, and of the undesirable traits, *dirty*, *stubborn*, and *dangerous*. The following sequence of events took place:

15 *Ss*: First Rated Mexicans—Next Were Frustrated—Then Rated Japanese
16 *Ss*: First Rated Japanese—Next Were Frustrated—Then Rated Mexicans

Since there was little difference between the two sub-groups the data were combined; the basic attitude measure was the number of favorable traits checked. Before frustration, a mean of 5.71 *favorable* traits were checked, as a consequence of frustration, 4.11. If the true differences were zero, a difference as great as this would be expected, i.e., would occur by chance, only three times in 100. Before frustration, 2.93 *unfavorable* traits were checked as characteristic of the two nationalities, as a consequence of frustration, 3.26. The difference between these two means is not significant. It is necessary, of course, to use a control group to determine changes that occur without the intervening frustration. Results from such a control group (N not specified) showed slight changes in the opposite direction from those above, i.e., a slight increase in number of favorable traits.

The authors assert on the basis of comments made by *Ss*, and from general observations during the experiment, that aggression was present. The data suggest that this aggression may have shown itself against objects (people) which had nothing to do with the frustration; hence, evidence for generalization or displacement of aggression.

Failure of paper and pencil tests to measure aggression. In an experiment by Sears, Hovland, and Miller (341), paid *Ss* were asked to report to the laboratory for an all-night study of the effects of fatigue on physiological processes. At the laboratory *Ss* were confined in a relatively bare room, were not allowed to smoke, were promised meals at various times but no meals came, and were, for various periods, required to sit still without any form of amusement. Such conditions prevailed throughout most of the night, and *Ss* were not allowed to sleep. Several paper and pencil tests were administered before the frustration period and again late in the night at a time when frustration appeared to be

near a peak. These tests involved the objective measurement of such factors as the humor of jokes, the rating of friends on certain traits, the amount of annoyance produced by common happenings, and so forth.

*E*s were unable to find any evidence of frustration reactions as measured by these pencil and paper tests. There was no sign of increases in amount of aggression, yet from observation, aggression was clearly evident in *S*s' behavior. It was *E*s' belief that two factors operated to prevent the aggression from showing itself in the tests: (1) the social situation, in which faculty members were in charge, may have produced marked inhibition of aggression in these paid *S*s; (2) the tests themselves may have acted as stimuli to induce a "test set," by which habitual patterns of behavior were reinstated, obscuring or replacing aggressive tendencies.

That the situation did produce some aggression is supported by the following incident summarized in another report of the same experiment:

> . . .one of the subjects produced two sheets of drawings in which violent aggression was represented in an unmistakable manner. Dismembered and disemboweled bodies were shown in various grotesque positions, some drowned, some hanging, some merely stabbed and bleeding, but all portraying shocking injury to the human body. Furthermore, when the creator of these pictures was asked by another subject, who the people represented in the drawings were, he replied, "Psychologists!" And his fellow sufferers were all obviously amused (81, p. 45).

It is the opinion of these *E*s that time samples or records of actual behavior in the frustration situation (as represented by the above incident) will in the long run be more valuable than paper and pencil tests as measuring instruments. The study itself suggests that college students will inhibit, or at worst, displace, tendencies of aggression toward faculty members. That is fortunate.

Measurement in Terms of Performance Changes

This method attempts to answer the question: "How does frustration affect performance on a standardized task?" The method may or may not be concerned simultaneously with deviations discussed thus far, e.g., aggression or withdrawal. Nevertheless, an

answer to the question is important for the better prediction of behavior.

Regression in children. A study by Barker, Dembo, and Lewin (10) is somewhat hybrid with regard to placement under a given method. The method of gathering the data illustrated continuous recording of behavior during frustration, but the major quantitative index concerned performance level.

The 30 pre-school children used in this study ranged from 28 to 61 months of age. Before starting the experiment *E*s spent 10 days at a pre-school getting acquainted with *S*s and taking part in their activities. No *S* was used unless he wanted to go to the experimental room. Each *S* served in two sessions. On the first, he was brought into the experimental room to determine his behavior under non-frustrating (control) conditions. Toys were available, and for 30 minutes the child played with these toys without *E*'s participating except as necessary to make the child feel secure. *E*, for the most part, sat at a desk and recorded a running account of the child's play. An observer behind a one-way screen independently recorded what happened. The observer and *E* alternated in their rôles from *S* to *S*.

When the child arrived for the second session he found that an opaque partition in the experimental room had been removed and that behind this partition was a group of new and exciting toys, including a large and elaborate doll house, a small lake with an island, lighthouse, beach, ducks, etc. The toys with which the child had played on the preceding day were scattered among the new toys. *S* immediately became engrossed in the new toys. After *E* was sure that the child had seen all the new toys and was thoroughly enjoying himself with them, he picked up the old toys (the ones used on the first day) and took them to the other end of the room (the section used on the first day) and placed them just as they had been previously. Then the child was asked to come and play with the old toys. After the child had left the new toys for the old ones, a wire-mesh screen was lowered to cut off the end of the room holding the new toys. *E*, at this point, said only: "You can play on this side now" (p. 60). The child could now see the new toys but could not get to them. This constituted the frustrating situation. *S* remained in this situation for 30 minutes, exactly the amount of time spent in the same room

during the first session. Deviant behavior was determined by comparing behavior on the first day with that on the second.

The basic measure used to determine change in performance was called "constructiveness of play." This is defined as "the degree of creativeness, elaborateness, or complexity of an activity" (p. 102). Obviously, a two-year-old child does not play in the same manner that a five-year-old child plays. The older child is more creative, drawings are more complete, and his imaginal processes in general are more elaborate. Thus, constructiveness of play, if quantified, should increase as chronological age increases. Furthermore, since mental age normally increases with chronological age, all three measures—constructiveness of play, chronological age, mental age—should be highly correlated. Finally, if regression occurs in this situation, it might be measured in terms of lower constructiveness of play in the second experimental period as compared with the first. This would indicate reversion to a form of behavior appropriate to an earlier age level, or *primitivation*, as it is called. The immediate problem, then, is to determine a reliable index of constructiveness of play.

In order to quantify the play index, units of behavior for each child were typed on cards. Note that these were units of behavior; not time units. Thus, if a child spent 100 seconds playing with a doll and then moved on to play with a truck, the play with the doll constituted a unit of behavior. The same would have been true had the child spent 200 second playing with the doll. Typed on the card for each unit of play was the merged account taken from the independently written records of the two *E*s. Three judges conferred on the behavior indicated on each card and ranked it along a seven-point constructiveness-of-play scale. Each rank order was assigned a numerical weight, with *two* being the lowest weight and *eight* the highest. These two extremes may be illustrated from the protocols:

Low constructiveness, numerically 2: "Sits on floor and takes truck and trailer in hand" (p. 102).

High constructiveness, numerically 8: "Detaches trailer, uses it as incline against ironing board. Runs truck up, carries it farther and farther, and lets it go. Looks to experimenter for approval, smiling. 'Did you see it? Now watch it.' Pushes truck across floor, big push. Hits *E*. 'See how fast it goes!' 'Chugs' it over to observer's window, looks underneath, 'Chugs' to table, to barrier, 205 seconds" (p. 104).

The constructiveness-of-play scale was shown to have high reliability, i.e., behavior units would be rated about the same at different times. Too, the scale was shown to correlate quite highly with mental age and chronological age. Hence, the instrument may be used to determine whether or not regression took place during the second experimental period. The evidence is positive. During the first experimental session the mean constructiveness of play for all *S*s was rated 4.99, whereas during the 30 minutes following the lowering of the partition the mean constructiveness of play in the same room with the same toys was 3.94, a difference which is highly significant statistically. Since constructiveness and mental age are highly correlated, the results could be expressed in terms of mental age units. The mean regression in terms of mental age was 17.3 months. Roughly speaking from the estimate of regression actually obtained, a frustrated five-year-old child would be about as constructive as a nonfrustrated three-and-one-half-year old.

It may be questioned whether or not this particular experimental design is adequate. No control group was used, and it is conceivable that the measured decrease in constructiveness of play would have taken place even had frustration not been introduced. For example, if a child is "forced" to play with the same toys for 60 minutes (two 30-minute periods), he might gradually become bored with them and his play become less constructive. A control group given the two 30-minute periods of play, without frustration before the second period, would have given a check on this possibility. In defense of the experiment as performed, it should be said that the data show constructiveness remained about constant during the first experimental session. Thus it could be argued that there is no regular change in constructiveness without frustration. However, such a change might have been observed during the second period had such a play period been given without prior frustration. Also it might be said that a second period of play without frustration might have shown a *rise* in constructiveness, in which case the data as given underestimate the amount of regression. The main point, however, is that for adequate quantification, a control group must be used as a base from which deviation in behavior of an experimental group can be measured. The published report contains many other references to frustra-

tion in terms of aggressive behavior, emotionality, and so forth, but this in no way demands that regression take place.

Frustration and mental age. The data above suggest that if we measured directly we would find a decrease in mental age as a consequence of a frustrating situation. This hypothesis has been put to test by Lantz (210).

Ss were school boys with a mean chronological age of nine and one-half years. All experimentation was done individually after proper rapport with each *S* had been established. First all *Ss* were given a part of one form (L) of the Stanford-Binet intelligence test. Then, one group was allowed to succeed at a game and another group was failed. Immediately after success or failure both groups were given another form (M) of the intelligence test. Data were also available on a third group which had taken both forms of the test with neither success nor failure intervening. Let us summarize the procedures for the three groups:

Group I (N = 107) Form L . . . Succeeded at Game . . . Form M
Group II (N = 105) Form L . . . . . Failed at Game . . . . . Form M
Group III (N = 108) Form L . . . . .Form M

TABLE 17

Effect of Success and Failure (Frustration) on Intelligence Test Scores

Group	*Form L*		*Form M*		*Mean Increase*
	M	S.D.	M	S.D.	
I	6.31	2.04	7.60	1.44	1.29
II	5.56	2.11	5.46	2.02	—.10
III	5.93	2.20	6.83	1.99	.90

Data from Lantz (210).

Table 17 shows the basic results. Let us examine these carefully. The total score which could be made on either test was 9. We note that the mean score for Group I is 6.31 on Form L whereas the mean for Group II is 5.56. The difference, .75, is significant statistically, i.e., if the true difference between the groups were zero, a difference this large would be expected only once in 100 times by chance. In short, the two groups were different in performance before the variable was introduced.

What could account for this difference? There are two possibilities: (1) the two groups were not random samples from the same population, or (2) *E* treated the two groups differently when administering the first test (Form L). The second interpretation is a possibility since other data show that the two groups were roughly equal on a full-fledged intelligence test. Why would they be equal on one intelligence test and not on the other, assuming the correlation to be high between the two? We cannot tell which interpretation is correct, and Lantz has not attempted to account for the discrepancies. In our evaluation of the data in Table 17 we must bear these initial differences in mind.

Group III is a control group showing how much change would be expected on the basis of practice alone in taking the two tests. The increase in performance from the first to the second test is evident. However, it is noted that Group I increases somewhat more than Group III, whereas Group II shows a decrease in performance on Form M as compared with Form L. Thus, in spite of the initial differences in performance on Form L, Lantz is able to conclude that scores on an intelligence test changed significantly as a consequence of her two conditions, success and failure. The depressing effect of failure seems to be greater than the enhancing effect of success.

E also had an independent observer rate *S*'s behavior on certain characteristics. In taking the second test, the failure group was rated as showing significantly lower motivation than the success group. In general, from the descriptions given, one obtains a picture of apathy in the failure group, although other characteristics such as anxiety, restlessness, and antagonism were observed.

Bulldozing responses and performance. There is probably considerable truth in the belief that under strong emotional stimulation (such as anger) rational solutions to problems occur with less frequency than during moderate emotional states. The procedure to be considered illustrates what may happen to performance in a frustrating situation where bulldozing activity and emotional excitement are also present. There are four experiments involved, two by Hamilton (135, 136), and two by Patrick (282, 283).

Hamilton's experiments were designed as comparative studies on the problem-solving ability of various organisms. Patrick was

able to reproduce Hamilton's results on humans and then in the major experiment added the frustration condition to determine what influence this had on problem solving.

The problem situation devised by Hamilton is represented in Fig. 34. Since Hamilton worked with humans, rats, cats, monkeys, dogs, and horses, the situation in terms of apparatus size and door-opening detail varied somewhat to fit the organism, but the basic problem remained the same for all. The essential features are four doors which are at an equal distance from the entrance to the enclosure. From our standpoint the problem is quite simple. *S* was brought into the runway and required to find exit from the enclosure. Only one of the doors was unlocked on each trial. The only "system" involved was that the same door would never be unlocked on two successive trials. Thus, if door one was unlocked on trial one, it would be locked on trial two. On trial two, one of the three other doors would be unlocked, which one being determined on chance basis. On trial three, the door unlocked on trial two would be locked, but one of the other three unlocked, and so on.

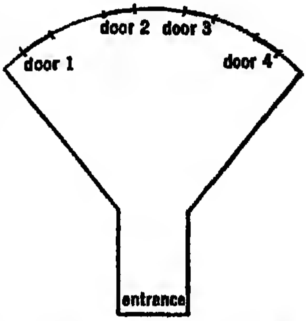


FIG.. 34. Problem situation devised by Hamilton (135) and also used by Patrick (282, 283).

As a result of his studies with various animals, Hamilton indicates five types of solutions used with some consistency:

Type A. *S* avoids door which was open on previous trial and goes in regular order to the other three, trying each only once. This is considered the best solution.

Type B. *S* tries all four doors once in irregular order. The door open on previous trial was not necessarily avoided.

Type C. *S* tries all four doors once in regular order. *S* does not avoid door which was open on previous trial.

Type D. *S* tries one door more than once but only after trying another door. Thus, door one is tried, then door two, then back to door one.

Type E. *S* tries same door several times before moving on to another one.

These solution types, from *A* through *E*, may be thought of as indicating the range from the most adaptable to the least adaptable behavior. Humans, on successive trials, exhibited a large proportion of *A* solutions, while gophers, for example, had a high proportion of *E*-type behavior.

Patrick worked exclusively with college students. As control behavior, he recorded the solution types used by 6 *S*s on 10 trials

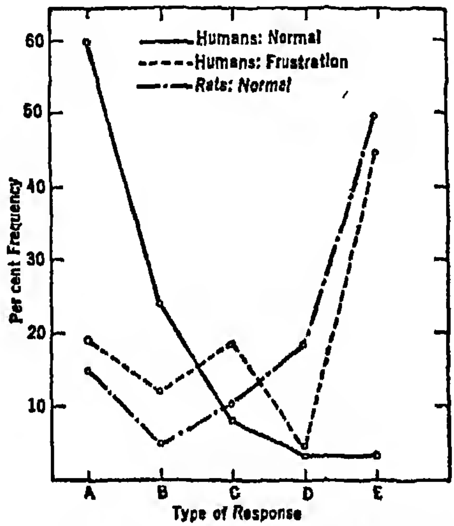


FIG. 35. The effect of frustration on problem-solving ability. Data from Patrick (283) and Hamilton (136).

a day for 10 days. Following this *S*s were given 100 more trials at the same rate but under extremely trying circumstances. *S* was dressed in a pair of coveralls. As soon as he stepped across the entrance of the enclosure he was shocked continually through the feet by a grid which covered the floor. Also, ice-cold showers were directed at him and a loud automobile horn was blown. Note that these conditions did not block the goal response as such, but did provide an extremely distracting and irritating state of affairs.

That the behavior of the students changed as a consequence of the unusual treatment is evidenced in Fig. 35. One curve shows the relative frequency of the five solution-types on the first 100

control trials, in which there was a total of 145 classifiable responses. Another curve shows the relative frequency under the exciting conditions of the second block of 100 trials, during which a total of 242 reactions could be classified. For comparative purposes, the solutions shown by Hamilton's rats are drawn. The data show that under strong stimulation the solution behavior of these college students was almost equivalent to that of white rats under normal conditions. The bulldozing response was clearly evident to the observer.

In a sense, Patrick's results may be viewed as a form of regression of the primitivation type (see p. 221). Hamilton put a six-month-old baby through this problem and Patrick's results show that college students' problem-solving under stress conditions was almost down to the level of the baby's performance (without stress).

From the studies reviewed thus far, and from other results not included here, e.g., Sears (336, 337), Williams (412), McClelland and Apicella (227), Alper (3), it appears that when the response to frustration includes strong emotional excitement we may expect a decrease in performance on at least certain types of tasks, and when response to frustration includes apathy as a result of continual failure we may again expect a decrease in performance level.

Experimental repression. Repression may be thought of as accelerated forgetting. It is believed to be brought on by anxiety built up by frustration of ego-motives. We shall devote a later chapter to the topic of forgetting, so for the present we need only demonstrate a method by which repression is measured.

One of the problems in an experiment by Rosenzweig (323) was to determine whether failure on a series of alleged intelligence tests enhanced the forgetting of those tests. On the assumption that failure on an intelligence test constitutes an interference with the satisfaction of an ego-motive, and hence arouses anxiety, repression might be expected. Two groups of 30 *Ss* each, one experimental and one control, worked at a series of jigsaw puzzles.

Ss in the control group were informed by *E* that he was constructing these puzzles and wanted to know if they worked out in practice. He attempted in every way to make it appear that

not the *S*s themselves but the puzzles were on trial. No indication of any kind was given the control group that ability to solve puzzles was a measure of intelligence. In short, Rosenzweig did not want to ego-involve his control *S*s.

The experimental group was instructed differently, as shown by the following excerpts:

> I am going to give you some jigsaw puzzles to do as an intelligence test so that you may be compared with the other persons taking the test. . . .
>
> Every puzzle counts as much as every other in the score. . . . If you do not solve any puzzle in the allotted time, I shall naturally be obliged to stop you. . . .Your work will be interpreted as representing the full extent of your ability, so do your best (p. 66).

Comments made by *S*s in the experimental group indicate that *E* was successful in making them believe ability to do jigsaw puzzles signified something. For example, one *S* reported: "My feeling was one of desperation. With each puzzle that I missed the thought forced itself upon me that I wasn't getting on well, that I had no business being at Harvard or in the field of Philosophy" (p. 66).

E, by controlling both the time allowed on each puzzle and the difficulty of the puzzles, caused *S*s in both groups to fail in completion of half the puzzles but aided them in completion of the other half. Each *S* worked on 18 puzzles of which 9 were completed and 9 were not. Immediately after *S* had worked all 18 puzzles, *E* asked him to write down as many names of the puzzles as possible. Each puzzle pictured some common object, so the naming was simple if *S* remembered.

The results, in terms of number of *S*s in each group recalling completed and incompleted puzzles, are shown in Table 18. Two facts are clear: (1) in the control group, more unfinished than finished tasks are recalled, and (2) in the experimental group more finished than unfinished tasks are recalled. Thus, the results for the two groups are in opposite directions. The differences in the recall of the two groups are highly significant statistically and may be clearly convincing as long as we assume that the two groups were equal at the outset of the experiment.

The difference between the two groups may be a function of two diverse phenomena. One, as observed in the control group,

is that unfinished tasks are better remembered than are finished. This phenomenon, called the *Zeigarnik effect* after its discoverer, has been verified on enough occasions to establish it. Its explanation is obscure. The second phenomenon, as indicated by the experimental group, appears to cónfirm the hypothesis on which the experiment was based. More finished tasks than unfinished tasks are remembered, a fact that is presumed to be indicative of repression, since failure to complete a puzzle produces frustration and anxiety, thus initiating a tendency to forget the humiliating experience. It is worthwhile to note that in one of a series of experiments by Rosenthal (321), very similar to the one reported here, the repressive effect was removed by hypnotic suggestion.

TABLE 18

Recall of Completed and Incompleted Tasks

Group	*Number Ss recalling more completed than incompleted tasks*	*Number Ss recalling more incompleted than completed tasks*	*Number Ss recalling equally well*
Control	7	19	4
Experimental	17	8	5

Data from Rosenzweig (323).

Hoarding in rats. The response investigated under this title is one we have not mentioned previously. Since this phenomenon can be shown to be a function of the blocking of a physiological motive, and since there is a rapidly growing body of literature on it, we should be acquainted with the basic techniques used. The apparatus, as designed by Hunt (172), consists of a home cage connected by a short alley leading to a food box. Hoarding behavior is indicated when the rat carries more food to the home cage than it eats. The amount of hoarding is measured by the number of food pellets carried but not eaten. Often the rat buries or hides the food pellets in some fashion.

Frustration, as preliminary to observing hoarding behavior, may be produced in two ways: (1) by putting the animal on a diet which maintains life but which leaves the animal hungry

(172); and (2) by putting food before a hungry rat but not allowing him to eat it, e.g., putting the food in a wire container which the rat cannot enter (229). Both methods produce a marked increase in the amount of hoarding as compared with control groups. The amount of hoarding is a function of several other stimulus variables in addition to those related to frustration (267), but clearly those conditions associated with frustration are highly significant.

There are numerous other studies concerned with performance changes and frustration, but we have covered a good sampling of methods and results. Whenever frustration is evidenced by aggression, withdrawal, or other identifiable mechanisms of adjustment, we may also expect performance changes. Most studies have shown a decrement in performance on a standardized task, but further analyses will very likely show that whether or not there is a decrement is dependent primarily upon (1) the type of task on which the performance is measured, and (2) the duration of the frustrating situation.

Measurement of Group Responses

One of the more recent trends in the area of social psychology is shown when *groups of Ss* are subjected to certain experimental conditions and the responses of the group as a whole analyzed. The methodological problems posed by these procedures are somewhat different from any we have yet faced.

As our illustrative problem we will consider an experiment by French (101). He wanted to determine whether there were differences in the reactions of organized and unorganized groups to a frustrating problem situation and to a fear situation. The fear situation was basically a frustrating situation since it involved escape from a locked room of a building in which a fire had ostensibly broken out. The locked doors blocked the goal-directed response of getting out.

Each group of *S*s consisted of six members. There were eight organized groups and eight unorganized groups. The latter consisted of Harvard University undergraduate students who had never seen each other before the start of the experiment. Five of the organized groups were basketball and football teams from the various college houses. The three other organized groups

were athletic teams from boys' clubs, their ages ranging from 12 to 14 years.

The experiment was carried out in an attic room, the floor plan of which is shown in Fig. 36. By reference to this floor plan you can follow the method of inducing the frustration. Note that there are no windows (though a skylight) and only one exit from the room. *Ss* (as a group) were brought into the room and seated around the table. They were told that the observers (five

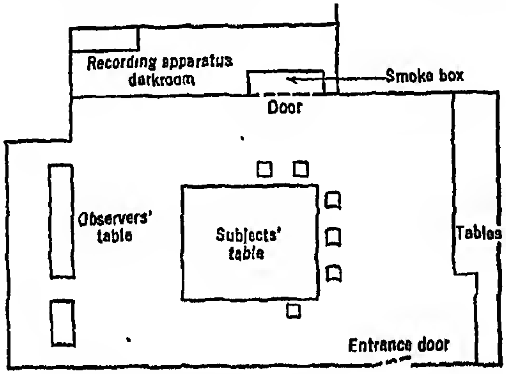


FIG. 36. Floor plan of situation used by French (101).

of them, seated at the table at the left end of the room) were to be ignored, since they were there merely to record the steps in the solution of the problems. *E* then showed the group three problems and read the instructions:

> The purpose of this problem-solving experiment is to discover *how fast the group can solve a problem.* Three problems are provided, but you are to solve *only one.* The group may choose which problem to work on, and it is permissible to shift to a different problem whenever the group desires. However, the whole group must work on the same problem at any given time. Thus the solution of *one problem in the shortest possible time* depends partly on knowing which problem to choose and when to shift to another problem (p. 237).

E informed the group that they would have only 45 minutes. The nature of the problems is not important; they were all difficult though not impossible to solve.

After 45 minutes *E* stopped the group's work on the problem

and took steps to initiate the fear situation. The observers were dismissed from the room. *E* informed *S*s that he had to leave early and requested that they remain and fill out a questionnaire after which they too could leave. As *E* left the room he surreptitiously locked the door. After a few minutes, *E*'s accomplice, hidden in the dark room, started the smoke box and soon wisps of smoke began to seep into the experimental room. The happenings from there on constitute part of the results. We must go back and consider the methods used in recording the behavior of *S*s.

Five observers were used. A concealed microphone connected to recording apparatus in the dark room provided a verbatim report of the first and last 10 minutes of the 45-minute period. Work on the first six groups was considered preliminary. From the preliminary work a check list of behavior characteristics was developed, the list consisting of such behavior as hostility, aggression, withdrawal, and friendliness. Various areas of behavior were divided among the observers so that each checked only one-fifth of the total behavior episodes. Each observer's checks overlapped that of another in at least one area in order that reliability of the judgments could be gauged. In addition to the check lists, each observer filled out four rating scales for each group. These scales concerned the amount of motivation, frustration, we-feeling (group unity), and amount of interdependence of the individuals. Ratings on these four factors were made every three minutes so that changes throughout the period could be noted.

In the fear situation the behavior was recorded by two observers watching through a one-way mirror inserted in the skylight, and by continuous recording through the microphone. After the fear situation, all *S*s wrote a description of how they felt and what they did when the smoke came into the room.

We shall consider the results from the problem-solving situation first. The observers' general ratings indicate that the organized groups showed significantly more frustration-produced behavior than did the unorganized groups. These over-all ratings were substantiated by the check-list results, in which it was shown that the organized groups exhibited a significantly greater number of aggressive responses. Both groups evidenced withdrawing

behavior, but no difference between the groups on this response was measurable. Other categories on the check list were shown to have low reliability, indicating that the observers were not well agreed upon the labelling or interpretation of certain responses. Such data are not scientifically useful. The total picture of the two kinds of groups indicates that the organized groups were more highly motivated, had more we-feeling and interdependence, and showed a greater number of responses to frustration in the problem situation.

The behavior of the groups in the fear situation was quantified only in terms of rank-ordering the groups for amount of fear shown. These ranks were obtained by examining the data from all sources. Only thirteen groups (seven organized and six unorganized) participated in this situation. Of these groups, the five ranked as showing the greatest amount of fear were all organized. These five groups attempted to break the door down and gain exit (to save doors *E* usually remained just outside and when it became obvious what was going to happen, he opened the door for *Ss*). The unorganized groups, on the other hand, were much more prone to view the smoke situation as a hoax and as a part of the experiment.

Interpretation of the differences in behavior of the organized and unorganized groups in the fear situation is difficult. French notes that the unorganized groups (university students) contained more men who had taken courses in psychology than did the organized groups. It might therefore seem reasonable that the unorganized groups would be less naïve toward the fire situation, hence, exhibit less fear. Differences in the fear situation may not be due at *all* to the unorganized-organized factor, nor to differences in the behavior of the groups in the problem situation.

The fact that French's results with regard to differences in behavior in the fear situation may be accounted for on the basis of factors other than the independent variable (organized-unorganized) raises a methodological problem which we should discuss.

Matched groups and social variables. For convenience, we will call any variable which influences the behavior of members of a group toward each other, or of one group toward another, a

social variable. Let us presume that we desire to determine the influence of different *kinds of leadership* on behavior of groups. *Kind of leadership* is a social variable by our definition. In planning such an experiment we would probably choose Design Method II, Matched Groups, although other methods have been tried (218). But on what factors are we to match groups when we seek to measure the influence of such a variable as leadership? The influence of a leader on a group is not a function of the individual *S*'s habit patterns alone, but is also thought to be a function of inter-*S* relationships. For example, a group consisting of three boys who were bitter rivals for a place on an athletic team might respond quite differently to a leader than the same three boys already well-established on the team. Clearly, in addition to the individual's abilities and capacities, inter-personal relationships in the group must be equalized for the various groups if matching is to be adequate. How, in view of all these considerations, are we to accomplish this?

To date, no generally accepted method of matching has evolved, largely because this is almost an untapped field of experimentation. One technique, that of Moreno (265), is worth reporting. This technique may be used to equate groups roughly on inter-personal relations. Let us say that a large group of *S*s (school children) has volunteered for our experiment, and that from this large group we wish to pick two groups of five each. All children in the large group are asked a series of questions to determine the following:

1. With whom they most like to work
2. With whom they least like to work
3. Who are their best friends
4. With whom are they least friendly
5. With whom they most often play
6. With whom they least often play, etc.

Knowing such facts for a fairly large group we would be able to select from the volunteers two groups of five each in which the number of likes and dislikes, friends and non-friends, playmates and non-playmates, leaders and non-leaders, and so forth, are about the same. We would attempt at the same time to equate the groups on variables such as intelligence, age, sex, or anything else we might think could differentially influence behavior in

the experiment. The procedure assumes that when a child is asked, for example, to name his best friend, the answer he gives is valid. Some check may be made on this by questioning teachers or observing the children on the playground or at work. Since a social variable, such as "kind of leadership," will influence the reactions of each member of the group to other members, the method of matching is that of equating groups on the same "task" which the variable is going to influence. This, as previously discussed, is the ideal procedure for matching groups.

STIMULUS VARIABLES INFLUENCING RESPONSES TO FRUSTRATION

When we ask what stimulus variables influence responses to frustration we are asking two questions: (1) What variables cause the *form* of the responses to vary; and (2) what variables cause the *amount* of a given form of response to vary? First we should know what conditions produce aggressive behavior; what conditions produce withdrawing behavior, and so forth. Then we should know what conditions produce variations in the amount of aggression or amount of withdrawing. It is very likely that the variables in the two cases will be the same; that is, those variables which are necessary to produce withdrawing will probably have dimensional relationships with amount of withdrawing.

We cannot at this stage of experimentation state the relationships between stimulus variables and forms of frustration responses. We are still in the stage of working out adequate measurements and classifications of responses. It is entirely conceivable that we are not yet able to discern and measure all significant responses to the blocking of goal-response. Furthermore, acceptable general empirical relationships between stimulus variables and responses to frustration cannot be stated until greater agreement is reached on the interpretation and classification of responses.

In the theoretical framework set up by the Yale group in *Frustration and Aggression* (81), aggression was assumed to be the primary response to blocked goal-response. The aggression, to be sure, could be displaced in devious ways, but it still

remained under the heading of aggression in the same general sense that the term has been used in this chapter. The authors were able to state three basic variables which determined the amount of aggression, and to adduce some evidence, experimental and otherwise, in support of the variables. In addition, certain variables were indicated which would determine the form of the aggression.

From the data we have covered we can see that aggression is a common response to frustration. It will be helpful, therefore, to indicate the variables which are presumed to cause its amount to vary. Also, it is probable that these variables will cause the amount of other frustration responses to vary, although no attempt will be made here to guess at these relationships. The three basic variables (assuming all other factors constant) proposed by the Yale group, and the relationships with aggression, follow. The terminology has been changed somewhat.

Strength of motivation. Holding the amount of interference constant, it is to be expected that the greater the motivation the greater will be the amount of aggression which will result.

Degree of interference. The assumed relationship is the greater the interference the greater the aggression. The McClelland-Apicella experiment, reviewed earlier, indicated this relationship. Quantitatively, the degree of interference might be varied in terms of duration of interference. Or, it might be determined by the level-of-aspiration technique and falsification of scores in terms of amount of falsification. That is, failure to attain an aspiration score by one point would be thought of as less interfering than failure to attain by five points.

Number of interferences. This may be considered as subsidiary to degree of interference. It is presumed that the greater the number of interferences the greater will be the aggression. The interferences must occur in fairly close temporal order if a summation effect is to be obtained, so it is evident that *temporal contiguity* of interferences is a sub-variable under this general variable. The experiments which employed various combinations of conditions for producing frustration were based on the assumption that total frustration would be greater for all conditions combined than for one taken alone. The principle as stated is easily supported by anecdotal evidence. For example, the man

who has a series of minor irritations at the office and then blows up at the slightest provocation at home is an American legend.

It should be understood that these three variables are stated in general terms. In the specific situation there are many ways to vary amount of motivation (Chapter VI), and presumably many ways will be found to vary amount of interference, and number of interferences.

Individual Differences

While it has not been the practice in this book to consider individual differences and their causes, it is especially pertinent to mention them here. People vary a great deal in their *frustration thresholds*, i.e., in how much interference with a goal-response can occur before deviant responses become evident. Probably as a result of differences in amount of experience with frustrating situations, a wide range of tolerances is developed. That experience with frustration is related to differences in frustration thresholds is indicated in a study performed by Keister and Updegraff (188). They were able to reduce responses to frustration in children by a 6-weeks' training program in which the difficulty of problems met gradually increased. By this method frustrations could be kept minor or short-lived.

Another important fact about individual differences concerns the mode of reacting to frustration situations. It is very likely that individuals develop somewhat stereotyped or habitual ways of responding to goal-blocking, some showing aggression in its various forms, some showing withdrawal, and so on. The ability to predict an individual's particular response in an experimental situation is not far advanced, although some evidence on subject variables is available, e.g., Lantz (210), and Zander (426).

CONCLUDING CONSIDERATIONS

In the material covered we have seen that *E*s have been most ingenious in devising situations to produce blocking of goal-responses. They have been equally resourceful in quantifying responses which do not lend themselves to traditional methods of measurement. One of these methods, that of the "human yardstick" (18), warrants a brief summary discussion and evaluation.

The human yardstick method. This method covers instances in which the observers *interpret* *S*'s behavior, thus in a loose sense serve as the measuring instrument. In the material we have covered the method has taken two distinct variations:

Variation I. The observer interprets the behavior at the time it occurs. He does this by giving it a label (e.g., aggression) or tallying it on a check list.

Variation II. The observer records everything that occurs as it occurs and interpretation is delayed until after the experiment.

Both variations have advantages and disadvantages. Variation I has the advantage of allowing an interpretation at the time the behavior happens. Certain nuances in *S*'s behavior may facilitate interpretation—a lifted eyebrow may tell an observer a great deal. In Variation II it may be virtually impossible for a group of observers to record all the little bits of behavior and yet these little bits may be important. Variation II, on the other hand, has the advantage that interpretation may be made at leisure and not under pressure of events as they happen in the experimental situation. A methodological study which determines the correlation between responses to frustration obtained by the two variations would be valuable.

Now how does the method satisfy requirements of scientific rigor? We know that the simplest form of measurement is counting. Counted data are often the only kind obtained in psychological experiments. We count the number of correct responses on each trial as *S* learns a list of words; we count the number of errors made as the rat learns the maze. The counts are made at the time *S* performs. Note that in Variation II above, behavior sequences are recorded and the counting is done *later, after it is agreed what should be counted.* Presumably in Variation I this agreement is reached before the experiment starts and the observers count or tally at the time of the behavior sequence.

The major difficulty in the human yardstick method concerns its reliability. Agreement among observers, as well as agreement of an observer with himself from time to time, must be high. Otherwise, the data are not scientifically useful. The facts are that observers can be trained to agree well on what behavior is to be called aggression, what behavior is to be called withdrawing, and so forth. That reliability is a problem is demonstrated in

experiments reviewed in which observers did not agrce on their interpretations of behavior.

There is no problem of validity in the results obtained by the human yardstick method. It can be seen that we have no ultimate criteria of aggressiveness or withdrawal to which we can appeal. The definition of the behavior categories is exclusively on the observer level; the categories are defined in terms of what an observer must perceive *S* doing or saying. It may be, of course, that future research will provide correlates of aggression other than those derived from the observer method. For example, research may give knowledge of rather consistent physiological changes which accompany frustration, though results to date have not been encouraging (e.g., Sherman and Jost, 352; Jost, 186; Davis, 77).

Our main point is that the human yardstick method, when reliable, is basically the same as all methods of experimentation. If changes in stimulus variables produce observer-noted changes in response, we are still relating stimulus and response variables. The only difference is that more interpretation may intervene between the response proper and the counting of that response than in the traditional experiment. The use of trained observers and a carcful delineation of behavior ranges which are to be counted in certain categories will reduce the hazards of this interpretive component.

A look ahead. What direction will be taken by future research? What are the major gaps in our knowledge? How far along are we toward an understanding of the behavior discussed in this chapter? The most pressing need is one of standardization, a standardization of several aspects of the experimental program.

1. First, we need a standardization of techniques to block goal-directed responses. This may not be too difficult to achieve, since certain techniques are being used more and more consistently, e.g., the level-of-aspiration technique.

2. The response side of the experiment causes greatest difficulty. There is as yet very little consistency in the methods of measuring responses to the blocking of goal-directed behavior. This leads to the third need of standardization.

3. Classification of responses. Since we know the responses only in terms of the methods by which we measure them, and since there is little standardization in measurement, inter-laboratory

communication is greatly impeded. If one *E* calls behavior measured in one fashion aggression, and another *E* calls behavior measured in a different fashion aggression, no quantitative comparison is meaningful.

In spite of these difficulties which need to be overcome, no one can deny that the last 15 years have seen great progress in this new field of laboratory experimentation.

SUMMARY

1. Frustration is a logical construct defined by relating stimulus situations in which there has been a blocking of or interference with goal-directed behavior to the deviant behavior which results from such a situation.

2. The following methods have been used to produce frustration:

   a. Interfering with primary motive
   b. Setting up physical obstructions
   c. Artificially producing failure, by:
      1. Setting unobtainable standards
      2. Falsifying scores
      3. Condemning *S* as inferior
      4. Combining these methods
   d. Creating conflict situations

3. Methods of response measurement and analysis:

   a. Analyzing verbal responses
   b. Analyzing verbal and motor responses combined
   c. Measuring by projective tests
   d. Measuring by questionnaire and ratings
   e. Measuring in terms of performance changes on standardized task
   f. Measuring group responses

4. Major stimulus variables influencing amount of aggression:

   a. Strength of motivation
   b. Degree of interference
   c. Number of interferences

5. The following unusual methodological problems were discussed:

   a. The basis for matching groups in a social situation
   b. The use of observers or judges as measuring instruments

CHAPTER VIII

Conflict

INTRODUCTION

Passengers on the Chicago elevated transit system are sometimes amused by the behavior of patrons as they leave the cars. Stations are so arranged that two exits from the platform to the street are available, usually about 50 to 60 feet apart. It has been observed that when a patron alights from the car midway between the exits, gyrations often occur in his behavior. He will start first toward one exit and then the other—a step in this direction, a step in that direction; or, perhaps a mere shuffle or weaving of the body will take place. Of course, the vacillation doesn't last very long—the "deadlock" is soon broken—but in simplified form the behavior illustrates the subject of this chapter. Two incompatible responses are evoked by this situation, and this defines a conflict situation.

Conflict and frustration. In the previous chapter it was indicated that conflict was a class name applied to a particular method for inducing frustration and measuring the consequences thereof. We will expect these consequences to be much the same as those outlined in the previous chapter: namely, such responses as withdrawing, aggression, and so on. In addition there are other measures peculiar to conflict. The over-all point of view is, however, that the present chapter is an extension of the study of frustration.

In many experiments on frustration we saw clear evidence of conflict in that the situation evoked incompatible responses. This was especially true when inhibitory responses tended to halt or drastically modify aggressive behavior. Let us be sure, then, we understand that conflict is one of the methods by which goal-directed behavior can be interfered with or blocked. From a sheer

procedural standpoint, conflict can be set off from frustration in most laboratory situations by the fact that *E* has control over the strength of the opposing response tendencies. This was not true in the studies outlined in the previous chapter.

The discriminal processes and conflict. When *S* is asked to judge which of two lights is brightest, and when the difference between the lights is slight, a miniature conflict may result. When *S* must express a preference between two objects which are very much alike, hesitations are evident. As seen in the early chapters concerned with such discriminations, relatively non-personal judgmental processes were accompanied by certain behavior (such as increased judgment time) which is on the same continuum with conflict behavior in the more personal or traumatic situation. The study of conflict may be closely linked in method and theory to the study of the discriminal processes as such. To produce the most acute conflict *E* attempts to design a situation in which a highly motivated *S* has a response tendency to each of two stimuli between which there is scarcely a discriminable difference. *Es* are limited in their experiments with humans since for obvious reasons severe traumatic situations cannot be produced in the laboratory. With animal *Ss*, however, rather violent conflicts may be induced.

Plan of the chapter. This chapter will be organized around the following major headings:

1. Generalized conflict situations
2. Specific methods used to produce conflict
3. Methods used to measure and analyze conflict responses
4. Major variables
5. Emotions and motivation

GENERALIZED CONFLICT SITUATIONS

Responses may be considered as either *approach responses* (toward a goal or incentive) or *avoidance responses* (away from a noxious situation). The most acute conflict occurs when two incompatible response tendencies are of equal strength. What are the possible situations which *E* may arrange to produce incompatible responses of near equal strength? There are four such situations: (1) the situation in which two approach responses

are evoked; (2) in which two avoidance tendencies are evoked; (3) in which both an approach and an avoidance response tendency are evoked; and (4) to take into account the fact that we often do not have just two response tendencies of one kind or another, a situation in which there are multiple response tend-

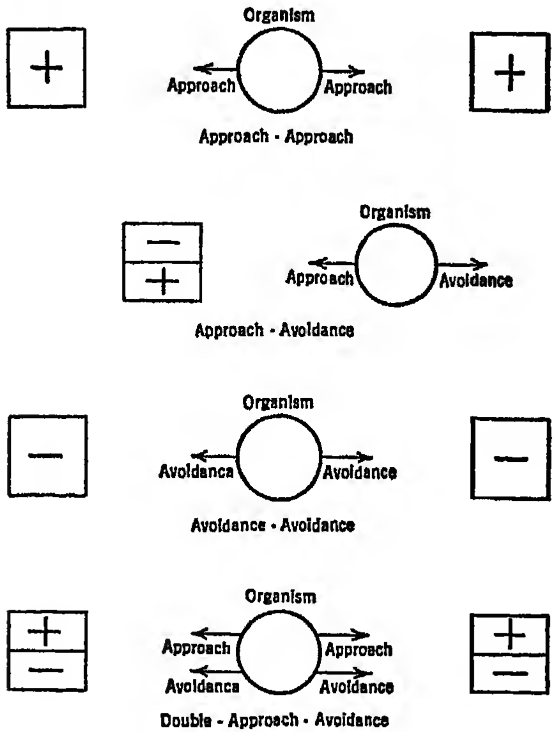


FIG. 37. Generalized conflict situations.

encies. The fourth situation will be called double-approach-avoidance. Lewin (216) has suggested the first three situations; Hovland and Sears (162) the fourth. All four conflict situations are shown in Fig. 37.

Approach-approach. In the approach-approach conflict the organism has two positive goal responses evoked at the same time, but by the nature of the situation both cannot be executed. A hungry child, offered a choice of one of two equally liked

foods, is in an approach-approach situation. The illustration of the elevated transit patron is another. Figure 37 shows the organism midway between two goal objects. Following Lewin (216), we may indicate the attracting goal object by a *plus*. Two incompatible tendencies are evoked simultaneously, and since the individual cannot go both ways at once, the situation may produce deviant responses. It is assumed, of course, that the motivational state of the organism is a part of the approach tendency.

The conflict situations in Fig. 37 are represented in spatial terms, i.e., in terms of distance to a goal, with *S* thought of as "locomoting" through space to get to or away from a situation. We must warn ourselves that many conflicts are not spatial in nature, but rather, psychological. By this it is meant that response tendencies are not necessarily opposed because their execution would mean locomoting in opposite directions in space. Opposition may exist because of differences described by other than the spatial dimension. Thus, the response tendencies of a person attempting to choose between two equally attractive job offers *would be thought of without reference to the spatial relations* of the job.

Much experimentation on conflict has been performed in terms of spatial distances between an organism and attracting and repelling stimulus situations. Equal distance in space from two goal-objects provides a fairly clear-cut method of observing reactions in terms of characteristics of locomotion. At present the assumption is that the basic principles established in the spatial situation will also hold in the nonspatial conflict situation.

Approach-avoidance. In approach-avoidance conflict a single situation or object has both positive and negative characteristics. Thus, both approach and avoidance tendencies are elicited simultaneously. If at any given point from the object the two tendencies are of about equal strength, the organism will theoretically be "suspended" motionless. The *minus* sign in the drawing indicates the repelling characteristic. Anecdotal illustrations are numerous: a child is attracted toward a vase on the table but he also has avoidance tendencies present as a consequence of punishment previously received for playing with vases; a soldier may have one response tendency to advance on the enemy and an opposed fear tendency to retreat. The term *ambivalence* is indicative of

an approach tendency and an avoidance tendency in one's reactions toward a person.

Avoidance-avoidance. In the avoidance-avoidance conflict the organism is diagrammed as being between two objects or situations, both of which elicit avoidance tendencies. Such a situation will usually lead to withdrawal (up or down in the diagram) if such withdrawal is possible. If not possible, serious behavior disturbances may occur when the avoidance tendencies are powerful. Indeed, there is reason to believe that suicide may sometimes be an outcome of an acute avoidance-avoidance situation in real life.

The student who does not like to attend class but who also does not like to be seen not attending is in an avoidance-avoidance situation. A girl, who is trying to avoid becoming an old maid, and who has received a proposal of marriage from someone she doesn't love, is in an avoidance-avoidance conflict. This type of conflict is implicitly assumed when we say that we choose the "lesser of two evils."

Double-approach-avoidance. This conflict situation is intended to make explicit the actual situation as it seems to occur in many instances: namely, that objects or situations do not evoke single approach nor single avoidance tendencies. Rather, several distinct tendencies of both kinds may be elicited by different aspects of the situation. We may, however, think of all approach tendencies for a given object or situation as summing into one "grand" approach tendency, and all avoidance tendencies summing into a grand avoidance tendency. With two objects or situations involved, the summation which takes place would result in a double-approach-avoidance conflict, i.e., a grand avoidance and a grand approach tendency for each of the two objects. For that matter, since there is no basis for limiting the number, we might have a triple-approach-avoidance situation, or a quadruple-approach-avoidance situation.

A hypothetical situation to illustrate this conflict may be seen in the man who is working in one position and is offered another. Each job has several attractive and negative features, such as pay, opportunities for advancement, prestige, and working conditions. If the attractive and repelling features for both jobs are equal when "summed," conflict will result.

An experimental illustration of the four situations. Before moving on to the various methods which have been used to produce conflict, let us examine a simple experiment in which all four situations have been studied in the same context. This particular experiment deals with motor conflict and may be conceived spatially in the manner implied by the diagrams in Fig. 37.

Hovland and Sears (162) constructed what may be called a *conflict board,* the essentials of which are shown in Fig. 38. The board used by Hovland and Sears was 5 inches square, but the writer has used boards the size of an ordinary sheet of paper with comparable results. The board is placed horizontally on a table or desk so that the two sets of lights are opposite *S*. *S* holds a pencil in the position marked "start," and is instructed that when the red light goes on he is to draw a line as quickly as possible to the other corner of the board (avoidance), and that when a green light is flashed on, he is to draw a line as quickly as possible *to* the green light (approach). *S*'s paper to draw on may be fed through on rollers, or individual sheets may be used, the former being most satisfactory. With the conflict board, Hovland and Sears set up the four conflict situations as follows:

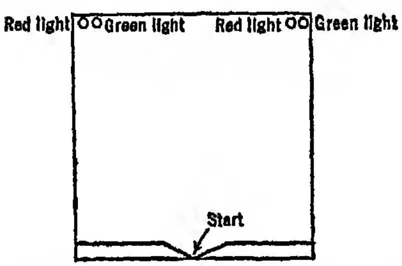


FIG. 38. Simplified version of conflict board for working out the four conflict situations (162).

Approach-Approach. *S* was given 20 trials, 10 each with the two green lights. The order of presenting the lights was randomly determined. On the 21st trial *E* flashed on both green lights simultaneously.

Approach-Avoidance. The red and green lights on *one* corner of the board were given 10 times each in random order. When a red light flashed, *S* drew a line to the opposite corner; when a green light flashed, he drew toward it. On the 21st trial, the test trial, both lights were flashed simultaneously.

Avoidance-Avoidance. The red lights at either corner were used in presenting 20 trials and again, on the 21st trial, *E* flashed both red lights simultaneously.

Double-Approach-Avoidance. All four lights were used and were flashed on singly in random order with the same general instructions prevailing.

Since this involves a somewhat more complicated situation than the previous three situations, *Es* allowed 80 practice trials and on the 81st trial all lights were flashed on at once.

What sorts of responses result on these critical test trials for the situations? First, *S* may be uninfluenced—he may draw a line directly to one side or the other. This would indicate no conflict. *S* may (and some did) respond to both lights by first drawing to one side and then quickly across to the other. A compromise response may occur in which *S* draws right down the middle of the board between the two lights. Some *Ss* blocked—no movement occurred, although tensing was evident. These deviant responses, deviant because they are different from those which occur in instances in which a single light is flashed, occur with relatively high frequency.

Motor conflict produced by these boards is highly predictable. The writer has used the boards for several class experiments, and even though *S* knows in general what is coming, the deviant responses will occur on a majority of test trials. This is especially true if the instructions for training emphasize a speed set. Furthermore, the deviant responses which occur in this simple situation have their rather striking parallels in the more traumatic conflict situations, which will be evident in material covered later in this chapter.

METHODS FOR INDUCING CONFLICT

Approach-Approach

There is some disagreement as to just what type of situation should produce an approach-approach conflict. On the one hand, we may present two equally attractive stimulus objects and inform *S* that he may have both, but locate the objects so that both cannot be obtained at the same time. *S* must go first to one and then to the other. Will his behavior be any different if a single object is presented? On the other hand, *E* may present the two goal-objects and indicate that only one may be obtained. The first situation is a conflict situation only because of temporal incompatibility, and there is no evidence that deviant responses occur in such a situation. The second is the one which will be used here as the prototype of an approach-approach situation.

There are two general methods of setting up an approach-

approach situation: (1) E may rely on response tendencies already a part of the organism's repertoire; or (2) the two approach tendencies may be built up independently. The first method would be illustrated by a situation in which E offers S two goal-objects known to be equally attractive. Thus, Godbeer (113) offered first- and second-grade boys the choice between two gumdrops of equal size and color. The second method would be illustrated by the conflict board situation in which S builds up approach tendencies to the green light (they are built up very rapidly). It would also be illustrated in an experiment by Klebanoff (200) in which rats were fed equally often at both ends of a straight-alley maze and then on the test trial put midway between the two food-boxes.

Goal-objects may be of different kinds and still be equally attractive. For example, in one experiment E determined how many tin soldiers were equal in attractiveness to a single gumdrop (113). This was done by a crude psychophysical procedure. In effect, E held a gumdrop and asked S how many tin soldiers he would *give for it. Three tin soldiers, let us say, might thus be found* equal to a single gumdrop. Then these two goal-objects could be used in an experiment on approach-approach conflict. There is some evidence from this study that when goal-objects differ in kind, though are the same in attractiveness, the amount of conflict is greater than if the two are of the same kind.

Approach-Avoidance

There are several methods by which this situation has been set up in the laboratory. We shall consider them separately.

Difficult discriminations. Pavlov (284) initiated this method and it has been used a great deal in attempts to produce "experimental neurosis" in animals. Pavlov used the conditioned response technique by which a formerly neutral or ineffective stimulus becomes effective in producing a response. He worked largely with salivary secretions in dogs. Meat powder or acid put into the mouth causes salivation. If a neutral stimulus, such as a bell or a light, is presented a number of times just before the salivation takes place, the neutral stimulus will itself come to evoke the salivary response.

To develop neurosis Pavlov used a sequence of events extending

over several weeks. In one of his experiments, the dogs were first trained to salivate to a luminous circle presented on a screen. This conditioned response was established without difficulty. Then a luminous ellipse, clearly different from the circle, was projected on the screen but was not followed by meat powder. The circle was always followed by meat powder; the ellipse was not. The animal had to learn *to* respond to the circle, and *not* to respond to the ellipse. The dog was able to develop this discrimination. Then Pavlov began to make the ellipse more and more like a circle, alternately presenting the circle and the ellipse. The dog continued making correct responses—not responding to the ellipse and responding to the circle—until the ellipse became so much like the circle that it apparently could no longer discriminate a difference in the forms. This occurred at a point where the vertical and horizontal axes were in a ratio of 9 to 8. It was after several weeks of alternate presentations of the circle and the ellipse at this ratio that the dog "broke down."

Much the same method had been used by many other investigators with a variety of animals. The basic procedure is that of training *S* to make an approach response to a stimulus and an avoidance response to a somewhat similar stimulus, and then gradually reducing the differences in similarity between the two so that both avoidance and approach responses are elicited by either stimulus.

Contradictory motivation. This is hardly an apt heading for the method, but in a sense it describes the situation. The method has usually consisted of depriving an animal of food or water so that motivation is strong, and then associating a fear or escape motive with the act of eating or drinking. Cook (64), for example, deprived rats of water and then "wired" a pan of water so that whenever the rat touched it he received a shock. If the strength of the shock became great enough, the animal refused to drink.

Masserman (250) taught cats to lift the lid of a box to get food and then during a series of trials he delivered an air blast just as the cat opened the lid. The blast appeared to be a fear stimulus. Thus, the animal had the conflicting motives of hunger and fear in the same situation.

Shocking on-going response. This procedure does not result in a clear-cut case of conflict, but because *E* has built up the two

opposing response tendencies we are including it here. The method is used with animals, commonly in a simple T-maze situation. Rats are first taught to go down one alley to receive food. After this habit is established, they are taught to go down the other alley to get food and thus abandon the first habit. At this point the animals have presumably developed an avoidance response to the first alley, and an approach response to the second. Then, on succeeding trials, the animals are strongly shocked just before they reach the choice point. *E* determines which of the two alleys is taken on these shocked trials and notes other behavior characteristics. If the animals take the first-learned path, regression is said to have taken place.

Avoidance-Avoidance

This method has not had many variants, but in one form it has been used a great deal, especially by N. R. F. Maier and his students at the University of Michigan. The basic apparatus is the Lashley jumping-stand (Fig. 39). Rats are placed on the pedestal opposite two small mounted doors or windows, the distance (six inches or more) from the pedestal to the windows being great enough so that the rat cannot "step across." By a gradual training procedure the rat can be taught to jump across the intervening space to the windows, the force of the jump opening the window and allowing access to food on the table behind the window. Then rats are taught a discrimination habit, let us say jumping to the window with a square on it and not jumping to the window with a circle. The window with the circle is consistently locked so that when the animal jumps against it, not only does he bump his nose but he also drops to the net below. Behavior of the rat suggests that this experience is not pleasant. In setting up the discrimination habit the right-left positions of the cards are randomized so that position habits are not learned, i.e., the rat doesn't learn to jump always to the left window, let us say, instead of to the window with the square design.

With the jumping-stand, as with other methods, *E* may gradually make the discrimination more difficult. However, when the similarity of the designs becomes so great that the animal cannot "tell a difference," he may refuse to jump even though under strong hunger motivation. Consequently, an additional "forcer" is

required. This has been provided in the form of an air blast or an electric grid on the pedestal. When such forcers are used the situation becomes an avoidance-avoidance situation. The card elicits avoidance tendencies but so also does the air blast or shock. We may expect deviant behavior to occur in such situations.

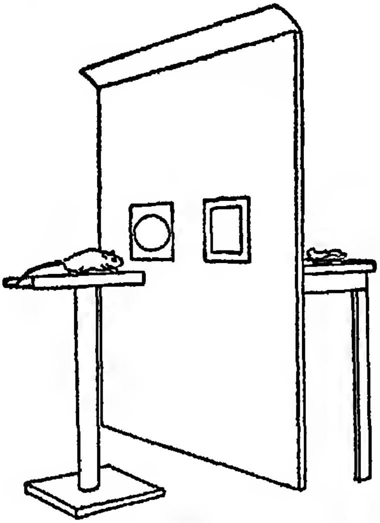


FIG. 39. Jumping stand. The net used to catch the rats following erroneous jumps is not shown.

In many instances *E* has made the problem situation insoluble rather than just a difficult discrimination. This has been done in several ways, all after the animal has learned to jump readily. *E* may lock both doors, may present only a negative door, or he may make neither of the cards consistently correct so that the animal can be only 50 per cent correct. Under certain conditions *E* has enclosed the pedestal on all sides (except the jumping side) and on the top. This restricts the range of orientation and tends to force the animal to orient toward the windows.

Double-Approach-Avoidance

Double-approach-avoidance conflict involves two objects or situations each of which has approach-avoidance characteristics. As yet, this situation has not been widely used in an experimental

setting. In the conflict board study (see p. 246) such a situation was devised by using both lights at both corners of the board.

One special method, used by Brown (40), is quite different from any of the methods discussed thus far. It makes use of a fundamental characteristic of behavior known as *generalization.* Let us define this term by experimental operations. *E* trains an animal to respond positively to a tone of 1000 cycles. After this response is learned it can be shown that the animal will also respond posi-

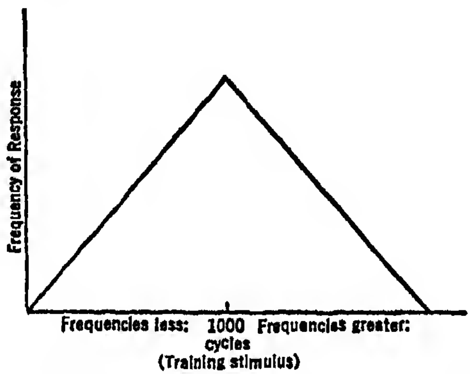


FIG. 40. Hypothetical gradient of generalization.

tively to tones of frequencies greater than 1000 cycles and less than 1000 cycles, even though these stimuli have not been used in training. The stimuli, though not identical, are thus functionally equivalent (they elicit the same responses). This equivalence, as measured by these operations, is called *generalization.* It can also be shown that the responses evoked by stimuli greater or less than 1000 cycles will occur with less frequency the greater the difference between the training tone and test tone. This suggests a *gradient of generalization,* in that the further the test stimulus is from the training stimulus, i.e., the less similar it is to the training stimulus, the less are the probabilities that the response will be evoked. This gradient is depicted in Fig. 40.

Now let us assume that if an animal is trained *not to respond* to a given stimulus there will also be a gradient of avoidance very similar to the gradient of approach. That is, stimuli greater or less than the training stimulus will produce avoidance behavior and

the farther away these test stimuli are along the dimension, the less frequent the avoidance response. Both the approach and avoidance gradients can be demonstrated empirically.

In a training situation such as a straight-alley maze, we now teach an animal to respond positively (approach the goal box) to a 1000-cycle tone and to avoid responding to a 500-cycle tone. These two gradients of generalization may overlap as depicted in

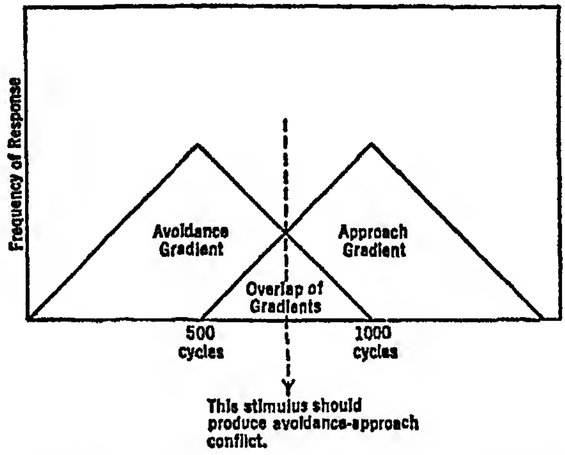


FIG. 41. Illustration of how overlap of approach and avoidance gradients may be used to produce approach-avoidance conflict.

Fig. 41. Now if we present a stimulus which is halfway between the two training stimuli (500 and 1000 cycles), at the point where the gradients overlap maximally, we should have produced an approach-avoidance conflict situation. Finally, if this problem were developed in a discrimination situation, and if on the test trials *two* intermediate stimuli were presented, a double-approach-avoidance situation should be created since we will have two situations of the kind depicted in Fig. 41.

The inducing of conflict on the basis of generalizations, whether approach-avoidance or double-approach-avoidance, makes certain assumptions concerning the distance of spread of the gradients of generalization and also the relative heights of the gradients. As drawn in Fig. 41, the gradients are of the same height and spread, indicating that the approach tendency is of the same strength as

the avoidance tendency. In actual practice the shape and spread can be determined empirically.

The situations outlined in this section will provide central points of departure in explaining methods of measurement in the next section. There are deviations from the methods discussed above, of course, but all major methods used to date for inducing conflict are represented in the various situations which we have considered.

METHODS OF RESPONSE MEASUREMENT AND ANALYSIS

The subdivisions of this section are formed in terms of various response measurements which have been used as indices of conflict. *E* may, of course, record several different responses in the same situation. A control group or condition is used to give base measurements with which conflict behavior can be compared, the control measurements being taken in a non-conflict situation. Like work on frustration, the work on conflict is still in the exploratory stage.

Latency

If two opposed response tendencies are simultaneously evoked, all logical considerations point to an increase in the time to respond as compared with time to respond when a single tendency is evoked. Very long latencies may be expected when opposed tendencies are of exactly equal strength. When two opposed response tendencies result in such a deadlock, *blocking* is said to have occurred. How long the latency of response must be before blocking is indicated is a function of the particular experimental situation and the latency which *E* defines as blocking in this situation. Let us consider two illustrations.

Godbeer (113) presented first- and second-grade boys with a choice between two equal-sized gumdrops. In the control condition *Ss* were instructed that they could have both gumdrops; in the experimental conditions *Ss* were told that they could choose only one. The apparatus was constructed so that two windows, 5 inches square and 30 inches apart, were at about arm's length from *S* as he was seated at a table. For a given trial, *E* opened a shutter exposing the candy in the windows. The opening of the shutter started an electric clock. *S* made a selection by moving a lever to

the window of his choice, thus hitting a botton, which stopped the clock and also delivered the candy. Thus, the latency was measured from the opening of the shutter to the hitting of the button under the window. When *S* was allowed to have only one piece of candy per trial, the mean latency of response was about 3.0 seconds; when *S* could have both pieces (control) the mean latency was about 1.75. The difference between the two means is highly significant.

As a second illustration we will consider an instance of exceptionally long latencies. In an experiment by Klee (201), rats were trained on the jumping stand to discriminate between a black circle on a white card and a white circle on a black card. One was used as a positive stimulus, the other negative. After this habit had been well established, Klee set up an insoluble situation by having one of the windows locked on each trial without regard to side or form on the window. Thus, if the rat jumped consistently to the black circle, he would be right only 50 per cent of the time. The rats were under 24-hour food deprivation at the start of the conflict series. The conditions of a control group were such that regardless of the window to which the animals jumped they always found it unlocked.

Latency measurements of the jumps were not taken systematically, but the extraordinary behavior of a few of the rats in the experimental group leaves no doubt that the conflict was severe. When first placed in the insoluble situation the rats at first jumped in accordance with their previously established discrimination habit. After a few trials they began to refuse to jump—the avoidance tendencies were stronger than the approach. For as long as 4 hours at a time the rats perched on the pedestal. Those animals that did not jump during the 4-hour period would not get fed so that by the following day they would be under 48-hour hunger deprivation. In some instances, animals went 72 hours without food before they jumped. For the rats which permanently refused to jump, it is presumed that the approach tendencies instigated by the hunger condition never became as strong or stronger than the avoidance tendencies produced by the insoluble situation and probable punishment resulting from jumping.

The jumping technique has been used by Finger (96) to show that as discrimination between stimulus figures becomes more

difficult (though soluble), latency of jumping increases. A correlated measure is the *force of the jump*. In general, as latency increases force of jump increases. It is as though the animal develops considerable tension during the latent period and this tension shows itself by the force of the jump.

Vicarious Trial and Error

During the latent period between the presentation of a stimulus situation and the time a choice is made, the organism is not inactive. We might suspect that there would be subtle behavior indicating conflict. When locomotion as such does not take place, but when there is evidence that the organism is actively engaged in "deliberation," vicarious (substitute) trial and error is said to be taking place. A rat at the choice point in a maze may be seen to look first down one alley and then another *as if* trying to determine which pathway to take.

Working with rats in an impossible discrimination situation, Brown (164) has shown that the number of head movements is directly related to the latency of response. The importance of such a finding is twofold: (1) head movements (swinging from one side to the other), indicative of vicarious trial and error, take place in a predictable fashion in a conflict situation; and (2) head movements might be substituted for latencies (or other measures) of conflict when latencies are difficult to measure.

In humans we might also expect the number of head movements to be indicative of conflict in a spatial situation. Such a measure has not been taken. However, eye-movements may substitute functionally for head movements, and the number of eye-movements has been shown to be directly related to the amount of conflict as obtained by other measures (113).

We may conclude that vicarious trial and error (as defined by the measurements indicated above) is found in conflict situations to a greater extent than in non-conflict situations. Hence, indices of vicarious trial and error may be used to infer the existence of conflict.

Withdrawal

Withdrawal may be *physical* withdrawal in the sense of locomoting through space to get away from the conflict-evoking

stimuli, or it may be *psychological,* in which case the organism "insulates" itself from these stimuli. We will consider an illustration of each type of withdrawing response.

Physical withdrawal. Brown (164) used rats as *S*s and a Y-type maze (Fig. 42). On all trials, whether a learning trial or a test trial, the rat was *not* placed at the base of the Y. Rather, it was placed immediately at the junction point as indicated in the diagram. If the rat were placed at this point, and if the stimulus situation elicited withdrawal responses, the animal could retreat into the base of the Y. Measurement of this tendency was effected in two ways: (1) by measuring the distance the rat retreated into the base, and (2) by an ingenious harness and pulley system whereby the strength with which the rat "pulled" to get back into the base was recorded. The harness was so constructed that forward movements, i.e., into either alley, were unimpeded. Retreat into the Y, however, was hindered by the light string connecting the harness with a system of springs and pulleys for measuring the strength of pull in grams. On the conflict test trials, Brown allowed the rat to pull for 5 seconds, then released the string so that the animal could retreat into the base. The distance retreated could thus be measured. Let us turn back to the methods used for inducing the conflict.

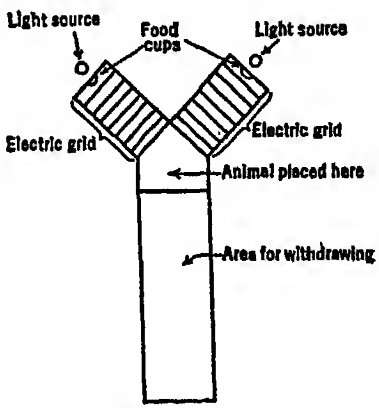


Fig. 42. Essentials of Y-type maze as used by Brown (164).

The two variable light sources (see Fig. 42) provided the basis for setting up a discrimination problem, and the grids were used for punishing incorrect responses. After preliminary training the rats were taught a discrimination habit. On each trial one alley had a bright light, the other a dim light. Half the rats were taught to approach the dim light (and receive food) and avoid the bright light. For the other half, the situation was reversed. The location of the stimuli was randomized between the two alleys on successive trials, to avoid position habits.

A total of 48 animals was used. Half the animals were shocked for going to the wrong alley, and half were merely deprived of food. Since the animals had already been divided into two groups on the basis of the positive stimulus used (bright or dim), this subdivision created four groups of 12 rats each.

All animals were trained until they made 12 successive correct responses. Training took place under 46-hour food deprivation. After learning, each rat was given three different test situations on one day, and three more test trials four days later. The only difference between the two test periods was that on the first day half the animals were under 46-hour food deprivation and the other half under 1-hour deprivation, whereas on the fourth day the situation was reversed.

The three test trials actually included one approach-approach situation, one avoidance-avoidance situation, and one double-approach-avoidance situation. To set up the approach-approach test situation, Brown flashed on two positive stimuli (one in each alley). Thus, if the animal had been trained to approach a dim light, on the approach-approach test trial a dim light appeared in each alley. The avoidance-avoidance situation was set up by putting on the two lights which the animals had learned to avoid. In the double-approach-avoidance situation, the overlap of generalization gradients was assumed and each of the two lights used had a brightness value halfway between the bright and dim light used during training. Under this situation each light might be expected to evoke both approach and avoidance tendencies, and since there were two lights, a double-approach-avoidance situation is presumed to have resulted.

Each rat was given all three test situations. We know that these test situations should not be given in the same order for all animals if we want to compare the conflict resulting from the three conditions. We also know that with three conditions it takes 6 *Ss* to counterbalance completely the order of conditions. Thus, for each of the four sub-groups of 12 rats each, two blocks of counterbalancing would be used. One block of 6 *Ss* had its first test trials under 46-hour deprivation; the other block under 1-hour deprivation. The deprivation time was reversed for the animals in the two blocks for the second set of test trials. This is a somewhat complicated design, though a very excellently planned one. We

will note that no two of the 48 rats went through exactly the same order of conditions throughout the experiment. Now, let us review the entire design, point by point, to see the purposes of the various sub-groups.

1. Half the rats were trained to approach the dim and avoid the bright light; half were trained under opposite conditions. By this procedure, any natural biases for approaching bright or dim lights should not prejudice the results.

2. The variable of shock or no shock for wrong responses is added by having half the *Ss* serve under one condition and half under the other. This gave four sub-groups of 12 animals each.

3. On test trials the variable of deprivation time was added. To counterbalance the order of testing, half the animals had the 46-hour condition first and the 1-hour condition second (four days later); for the other half, the conditions were reversed. Since each sub-group of 12 was split in half, this gives us eight groups of six animals each.

4. Finally, since the order of conditions for each test day might have influenced the results, the sequence of the three tests for each rat was different, for with the three test trials (approach-approach, avoidance-avoidance, and double-approach-avoidance) 6 rats would provide complete counterbalancing.

From this analysis we could see that a single rat might have the following conditions: an approach toward dim light; shock for errors; 46-hour deprivation on first test; 1-hour deprivation on second test, and test in the order of avoidance-avoidance, approach-approach, and double-approach-avoidance. No other rat followed exactly the same sequence throughout.

Both measures of withdrawing showed about the same results, so that one could be substituted for the other in this situation. We shall, therefore, present only results showing withdrawing in terms of amount of mean maximum pull exerted in "trying to get back" into the base of the Y. Graphical representation is given in Fig. 43. As we might expect, the avoidance-avoidance situation produced the greatest amount of withdrawal under all conditions. Also, as we might expect, the approach-approach situation instigated only slight withdrawing behavior. This is in line with other findings. The slight evidence for withdrawal under the approach-approach situation is due, Brown thinks, to avoidance

tendencies produced by the spread of the avoidance gradient to the positive stimulus.

The results show the importance of deprivation time as a variable. Greater deprivation time should increase the strength of the approach tendency and thus reduce the amount of withdrawal. This is clearly shown. Shock is shown to be influential in increasing avoidance tendencies, since with deprivation time held constant, greater withdrawal is shown in the rats which had been

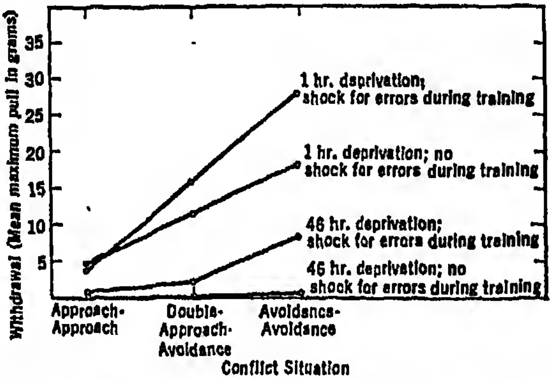


FIG. 43. Withdrawal behavior as a function of conflict situation. Adapted from Brown (164).

shocked for errors during learning. Comparatively, however, it can be seen that shock for errors does not add as much to the withdrawing tendency as 46-hour deprivation of food adds to the approach tendency. Increase in the strength of the shock would undoubtedly change this relationship. Or, it is possible that by reducing deprivation time to, let us say, 20 to 30 hours, the influence of the shock would just balance the approach tendencies produced by deprivation.

Psychological withdrawal. Suppose a child does not like to wash his hands and face and yet is punished or scolded when he doesn't. The child is told by the parent to "go wash your hands and face." This sets up an avoidance-avoidance situation. The child is not likely to leave the house—to withdraw physically—although such a response would not be unheard of. Rather, the child on the way to the bathroom may get side-tracked or

thoroughly engrossed in some other task *as if* he had forgotten his mission. This behavior suggests a form of psychological withdrawal.

Such considerations indicate that if we set up an avoidance-avoidance situation we might measure amount of withdrawing in terms of the number of extraneous and irrelevant activities in which *S* engages. Without detailing the experiments, we may say that this expectation has been confirmed by French (100) using college students as *S*s, and also by Sears with high-school students (337). The quantification consisted in both cases of measuring how long *S* spent on extraneous activities not pertinent to the accomplishment of a task toward which there were conflicting tendencies. Psychological withdrawal, as measured by time spent on extraneous activities must therefore be added as an index of conflict. Such withdrawal will occur only in instances in which the avoidance-avoidance situation exists, or in which the avoidance tendencies in the avoidance-approach or double-approach-avoidance situations are stronger than the approach.

Compromise Responses

On test trials in the conflict board situation, it was observed that some *S*s drew lines down the middle of the board—halfway between the two corners. This illustrates a compromise response since both response tendencies showed themselves in a single action. Similar results have been observed in other investigations (340, 355). Measurement of the compromise response may be made in terms of the amount of deviation from the non-conflict movement.

Compromise responses have been observed in animals but we have no instance in which these have been systematically measured. On the jumping stand, animals are sometimes reported to jump abortively when faced with an insoluble problem. Under normal circumstances the animal will hit the card or window with his nose and front paws. On the abortive jump it hits the card with the side of the body as if there were no expectation that it would gain entrance.

Cook (64) describes certain behavior in rats in an approach-avoidance situation which appears to be a form of compromise. Rats gradually approached the stimulus situation and then

suddenly jumped to the side, or jumped up on the side of the cage. The response took the animal no closer to the stimulus and no farther away from it.

Except in the simple spatial situation, such as illustrated by the conflict board, no quantification of compromise responses has occurred.

Seizures

As indicated previously, when a rat placed on the jumping stand in an insoluble problem situation refuses to jump, forcers are necessary. It was during an experiment in which an air forcer was being used that Maier first recorded extremely deviant behavior in rats (239). Following the air blasts some of the rats jumped from the pedestal, net, or feeding platform and ran wildly in a circular path about the room. Some ran headlong into walls, chairs, or table legs. After a short period of this violent running, convulsive-like behavior appeared. Body tremors were noted followed by a state of inertness in which *E* could "mold" the animal into various postures. Then there was a gradual return to normality. It appeared as if the avoidance-avoidance conflict situation had produced extremely deviant behavior.

Subsequent research by Maier and others have shown that it is not the conflict situation *as such* which can account for the deviant behavior. It now appears that a major factor in producing the seizures is the high frequency sound waves present in the air blasts. For example, seizures can be induced in some rats while in home cages merely by jangling keys. Since there is no reason to believe that conflict is present in such instances, the cause must be a susceptibility in some rats to the sound waves produced by the keys. How these sounds operate to produce the seizures is unknown, although a large body of literature has already been built up showing that numerous variables, such as drugs, temperature, and so forth, will influence the susceptibility of the rats to seizures. Finger has reviewed these variables as known through 1946 (97). Morgan and Waldman have called the seizures, induced by auditory stimulation, *audiogenic seizures* (268), and this is the term now in common use.

It is known that certain conditions will lower the audiogenic seizure threshold, thus making the animals more prone to exhibit

the deviant behavior. One of the psychological variables which is important in lowering the threshold is the avoidance-avoidance conflict situation. This is demonstrated in an experiment by Maier and Longhurst (242).

The experimental group of animals was trained on the jumping stand to make a discrimination between a black and a white circle, with the black circle being the positive stimulus. The habit was considered learned when 10 errorless responses were made on each of three consecutive days. At no time during training was an air blast used. Following this training the conflict situation was introduced by presenting a screen with only one window. Half of the time the card in the window was the positive stimulus and half the time it was the negative. If the animal did not jump during a 30-second period when either card was presented a blast of air of moderate intensity was applied. Previously, the intensity of the blast had been shown to be below the seizure threshold for most of the rats.

The control group was comparable in every way to the experimental group except that its members received no preliminary training on the cards. These animals were simply placed on the pedestal, and after 30 seconds were exposed to a blast of air. The animals did not jump from the blast of air since they had no training in jumping and the air blast was not strong enough to force a jump from the stand in the unfamiliar situation. Thus, this group was exposed to the same air blasts as the experimental group but they cannot be considered to be in the same conflict situation as was the experimental group.

TABLE 19

Influence of Conflict in Decreasing the Threshold of Audiogenic Seizures

Group	*Frequency of Seizures*			*Number Rats Showing Seizures*	
	Trials	Seizures	Per cent	Number	Per cent
Experimental (N = 37)	296	60	20.3	14	37.8
Control (N = 44)	352	27	7.6	4	9.1

Data from Maier and Longhurst (242).

The basic results are shown in Table 19, in terms of number of seizures and number of animals having seizures. Ten trials a day, half with the positive card and half with the negative card, constituted one test day, and the experiment ran for eight such test days. The data clearly indicate that some factor is added in the experimental group which increases the frequency of seizures over that shown by the control group. It is Maier's belief that this feature is the conflict produced by the insoluble problem situation. Thus, frequency of seizure becomes an index of conflict when compared with frequency in a non-conflict situation.

Performance Changes on Standard Task

As in the case of the study of frustration induced by methods other than conflict, performance on a given task may be expected to be influenced by conflict situations. Three different approaches to this problem will be studied.

Regression. We are already familiar with the concept of regression as indicating a reversion to an earlier form of behavior. In the previous chapter an instance of experimentally produced regression in children was discussed. At this point we wish to review the animal studies. The basic pattern of several experiments has been the same. Habit *A* is acquired first. Following this, an opposed habit, *B*, is learned to the same situation, thus requiring habit *A* to be abandoned. On test trials *E* gives the animal a strong electric shock and observes which behavior sequence follows, that is, whether the animal executes Habit *A* or *B*. If the shock causes Habit *A* to be reinstated, regression is said to have occurred since the animal is now "going back" to an earlier form of behavior. We need to consider three studies to give us the essential methods.

1. *The Mowrer Study.* Mowrer (274) constructed a small box with the floor entirely covered by a steel grid through which shock of controlled variable intensities could be administered. At one end of the box a small lever projected from the wall. In a normal training series the rat was placed in the box and the intensity of the shock gradually increased. The physical activity of the rat was directly correlated with the intensity of the shock and sooner or later the animal would (by chance) hit the lever in the box, shutting off the current in the grid. Immediately,

however, the current would start to build up and again the rat was required to hit the lever to escape the shock. Learning in such a situation is precipitate. The very wise rat might deposit himself on the lever and stay there.

In his demonstration of regression, Mowrer used an experimental and a control group. First, the experimental group was placed in the box with the lever withdrawn. The characteristic response shown by these rats was that of sitting on their hind legs which, it is presumed, diminishes the intensity of the shock. This habit, sitting on hind legs, may be designated Habit *A*. Habit *B* consisted of learning the lever-pressing response to escape the shock. The experimental group was able to accomplish this although it took considerable prodding to get the rats to abandon the sitting habit. A control group learned Habit *B* only. Thus we have:

Experimental: Habit *A* (sitting) . . . . Habit *B* (lever-pressing)
Control: . . . . Habit *B* (lever-pressing)

In order to induce a conflict situation on test trials, Mowrer wired the lever so that a shock would be received through it when the rat touched it. Thus, animals could proceed with Habit *B* but would have to *endure* a slight shock in order to *escape* a much greater shock to the feet. The results show that animals in the control group continued to execute Habit *B* for 10 trials, i.e., they "took" the shock through the lever. Of the five animals in the experimental group, however, four reverted to Habit *A* at the moment they felt the shock from the lever; that is, as soon as these animals felt the shock they sat up on their hind legs. The difference in the behavior of the control and experimental animals provides the demonstration of regression.

2. *A T-maze Experiment.* Several *E*s have studied regression with the use of a simple T-maze or modification thereof. The earliest study was performed by Hamilton and Krechevsky (137) and will be used here as illustrative of the general methods and results.

In this study the T-maze was modified so that one arm of the T was twice as long as the other. Habit *A* consisted of making a left turn into the short alley. If food is present at the end of both arms, rats will eventually come to take the shorter path to

food. The animals were given 10 free runs a day until 90 per cent of their last 50 runs were to the shorter alley, hence, a left turn. For establishing Habit *B* the maze was turned over so that now the "correct" path—the short path—was entered by a right turn. The right turn would thus be an antagonistic response to the left turn defining Habit *A*. Enough trials were given on Habit *B* so that it was clear Habit *A* was breaking down—the left turn was being avoided.

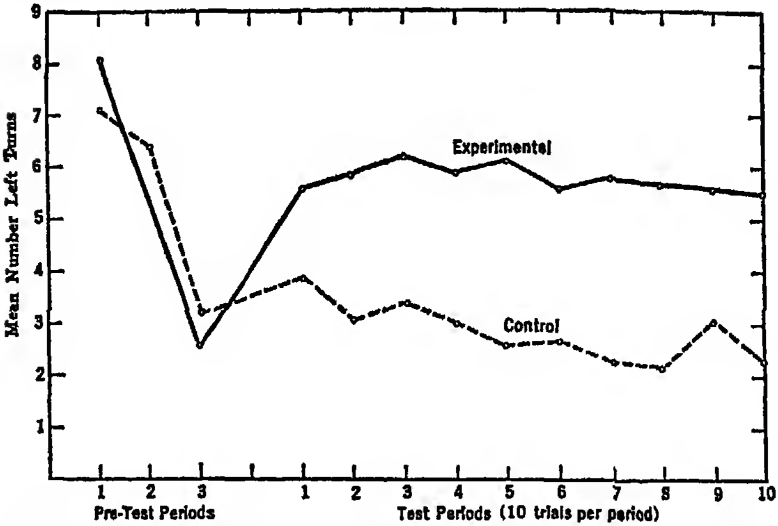


FIG. 44. Regression in rats. The experimental group was shocked at the choice point on all test trials. Data from Hamilton and Krechevsky (137).

The test periods consisted in shocking the rat just before it reached the choice point at the junction of the two arms. If regression occurs, the number of animals following Habit *A* should be greater than the number of a control group which had no shock. We would expect the control group to continue following Habit *B*. The results are shown in Fig. 44. Also shown are the last three training periods on Habit *B*, demonstrating that before the introduction of the shock the two groups were performing equally well on this habit. After the introduction of the shock, however, the experimental group shows a decided increase in the frequency of *A* responses. The number of *A*

responses remains relatively constant for a total of 100 trials (10 test periods). The difference in the behavior of the two groups must be attributed to the shocks given the experimental group.

3. *A Multiple-Choice Situation.* Kleemeier (202), after studying the methods and results of several investigations in the T-maze, suspected that one significant interpretation had been overlooked. When a two-choice situation is used, the influence of the shock might not be that of precipitating a return to an earlier learned habit as such, but rather one of bringing about an avoidance response to the continuation of the *B* habit. If the shock produces an avoidance response to *B*, where else can the animal go but to *A*? If this is true, Kleemeier argued, shocking a rat at the junction of a multiple-alley situation would show that in the cases of those rats not continuing Habit *B*, the choices would be about equally distributed among the other alleys. This would be evidence that the animals were not "going back" to Habit *A*, but rather, were avoiding *B*.

To test this hypothesis, Kleemeier constructed a maze with four arms leading off from the junction instead of the two in the simple-T. An animal was trained to go down first one of the four alleys (Habit *A*) and then down a different alley (Habit *B*), food being the incentive for both habits. All possible combinations of alleys were used for the different sub-groups of animals. One group would acquire *A* in Alley I, and *B* in Alley II; another group would learn *A* in III, and *B* in I, etc., so that alley bias should not prejudice the results. With four alleys there are six combinations of alleys taken two at a time. If we reverse the order of alleys for each of the six combinations the total number of sub-groups needed is 12. In this experiment, 60 rats were divided into 12 sub-groups of five each to effect complete counterbalancing of alleys taken two at a time.

After training, 50 trials were given with shock delivered just before the animal reached the junction point. Since there were 60 rats, a total of 3000 trials was involved. The results, as given in Table 20, show that about 50 per cent of the animals continued with the *B* habit. This figure is roughly equivalent to those found in several simple-T maze studies. To confirm Kleemeier's hypothesis that so-called regression responses may only be avoidance responses to the *B* alley, we would expect the remaining 50 per

cent of the runs to be almost equally distributed among the other three alleys, or about 17 per cent in each. This is not the case. On 27 per cent of the trials the animals took the Habit *A* alley, whereas they took the remaining two alleys about 12 per cent of the trials. There is thus some evidence for regression; training on a given alley determines to some extent the response which will be made when an on-going habit learned later is shocked. It is clear, however, that the evidence for regression is not strong, and this experiment shows the importance of the availability of other alleys.

TABLE 20

REGRESSION IN A FOUR-CHOICE MAZE

	Habit A	*Habit B*	*Digression*
Number Choices	805	1478	717
Percentage	26.83	49.27	23.90

[a] Habit *A* would be considered a regression response; Habit *B* a continuation of the response most recently learned, and the digression responses those cases in which rats took one of the two other alleys available. Data from Kleemeier (202).

At present the empirical facts concerning regression show that when an on-going behavior is disrupted the animal is likely to return to an earlier learned habit with greater than chance frequency. From several experiments we know that the amount of regression is a function of manipulable variables. For example, shock must come in the maze and not outside the maze—the shock effects are specific to the situation and the on-going performance of Habit *B* (330); regression may be evident at the first choice point of a two-unit maze but not at the second (295); the stronger the *B* habit the less the likelihood that regression will occur (249). One variable, intensity of shock, has not been manipulated. Differences in amounts of regression found among various experiments may well be a function of this factor.* So far as we can tell at present, the interpretation of regression, like the more

*One may wonder why the rat continues to run on successive trials when it gets shocked on each of these trials as in the Hamilton-Krechevsky experiment. An adequate explanation has been offered by Farber (93, Footnote, p. 115). The complexity of the explanation precludes its discussion here.

clear-cut conflict phenomena, must be based on learning principles. It is presumed that regression occurs when the avoidance of the *B* habit produced by the shock is stronger than the avoidance of the *A* habit originally set up when *B* was being acquired.

Fixations. Under the influence of strong shock, rats may develop extremely stereotyped responses. These responses are exceedingly resistant to change and will persist for many trials after the shock has been removed. Inasmuch as they are not appropriate to the situation after the shock has been eliminated, they may be considered a form of deviant behavior.

While most *E*s have demonstrated fixations by the use of shock, Klee's work (201) makes it clear that other forms of severe punishment may result in fixations. Klee first trained rats on a jumping stand discrimination problem. After training, the problem was made insoluble and the rats were forced to jump by air blasts or by very long periods of food deprivation. Klee observed that a rat would develop a characteristic response to this insoluble situation; for example, it always jumped to the left window. When the problem was then made soluble the characteristic response developed by a rat would persist over scores of trials, and, as compared with control rats, the emergence of an adaptive or correct response was long delayed. The difference in behavior of the control animals (not having had the punishment) and the experimental animals defines fixation.

Fixations are not to be thought of as responses differing in kind from an ordinary learned response. Farber's (93) theoretical and experimental analysis of fixations makes it appear very likely that these responses are produced by principles which are no different from those producing normal learning. We will not consider this matter further at this point since it would take us far afield into theories of learning, a topic we are reserving for a later chapter.

It is improbable that human punishment situations as used in the laboratory are comparable to those used for producing fixations in rats. Fixation by definition represents low variability of response, and with humans the usual effect of punishment is to *increase* variability (337). Fixations comparable to those produced in rats have not been produced experimentally in humans. However, the persistence of certain human habits in spite of

their non-adjustive nature makes it likely that such phenomena as fixations are not restricted to animal behavior.

Discrimination breakdowns. When a conflict is induced by pushing a discrimination problem to the point where the animal is no longer able to differentiate, it has been noted that the animal may lose its ability to make even much easier and well-established discriminations. This was first noted by Pavlov and has been verified by many *E*s. There have been no systematic quantitative reports on how the amount of loss of discrimination ability is related to such factors as length of training in the difficult discrimination situation, the amount of conflict as inferred from other indices, and so forth. However, the frequency with which the phenomenon of discrimination breakdown has been reported leaves no room for questioning its authenticity.

It is conceivable that we might present human *S*s with a discrimination situation which gradually becomes more difficult. In the kinds of situations which we might set up in the laboratory, however, *S*s would probably just shrug their shoulders and indicate they don't observe a difference. However, if some form of punishment were added, deviant behavior might result. In one experiment (99) two visual stimuli were presented to *S* with instructions to indicate which was the brighter of the two. *S* was given 10 seconds in which to respond, or else a shock was given at the end of the 10-second period. Moreover, if *S* named the wrong stimulus he was shocked. Initially, the two stimuli were easily discriminated—there was a very noticeable difference in brightness value. The difference between the stimuli was then gradually narrowed until *S* made three consecutive errors; then the descending order was repeated. Among the findings was the fact that as the number of series increased, *S* had more and more difficulty with the easier discriminations—those readily distinguished on preceding trials. This was likened to the breakdown of discrimination in animals.

It is difficult to determine whether the finding of decreased discrimination ability in this experiment represents a true state of affairs or is an artifact of the situation. *S* might soon discover that for any given series he was going to get three consecutive shocks sooner or later, and that when near the threshold he might actually get many more before a new series was started.

Thus, near the threshold he might make an error and get shocked; make a correct response and receive no shock; make two incorrect responses and get shocked for each, and so forth, for a considerable period of time before he actually made three consecutive errors and could start over again with stimuli which were easily discriminable. If we grant that *S* soon discovered that by voluntarily making three consecutive errors somewhat early in the series he could start a new series, this could account for what appeared to be a reduction in discriminative ability.

General Observations

Finally, in addition to the measures of conflict thus far discussed, we should mention that most *E*s record their general observations of the behavior of *S*s. Most of the observations are used for inferring the existence of emotional disturbances in the organism. Among animals, vocalization is a common response to a stress situation. Rats will squeal and struggle; dogs will bark, whine, and bite at the apparatus. Other animals show similar behavior. In certain instances (64, 219) animals seem to become hypersensitive to touch. Merely brushing the hand across the rat's leg may bring loud squealing and struggling. Certain other indices of upset have been used, such as amount or frequency of defecation and urination. The observations are sometimes reported in terms of percentage of animals showing the various kinds of behavior.

BASIC VARIABLES IN THE CONFLICT SITUATION

From the material we have covered, and from a few additional considerations, we will be able to point out certain manipulable variables which cause the amount of conflict to vary. In some cases the data are fairly clear-cut; in others it will be necessary merely to suggest the probable variables and the relationships which obtain.

In any conflict situation the most important single factor which will determine the amount of conflict is the *relative response strengths* of the opposed tendencies. Unfortunately, however, there are several conditions which *E* may manipulate to change relative response strengths so that in actual practice a

series of sub-variables will have to be indicated as those actually manipulated.

It is presumed that maximum conflict will take place when two opposed response tendencies are exactly equal in strength. It is doubtful that such a pure situation can be devised, but approximations to the condition are possible. Sears and Hovland (340) have found that if two incompatible movement responses are given an equal number of trials, greater blocking will occur on a test trial than if an unequal number of trials are given. The same general principle has been derived in verbal learning situations where two different responses to the same stimulus have been learned (383). There seems to be no reason to doubt the general principle that the more nearly equal are the opposed response tendencies the greater the conflict. In relatively simple situations at least, maximum conflict is to be inferred when blocking takes place.

It is reasonable to suspect that as the incompatible responses become more and more unequal in strength, there will be a succession of response-types, from oscillation or vicarious trial and error, to the point where immediate approach or immediate avoidance takes place (no measurable conflict). Thus, we might predict that different forms of conflict (along a continuum) will take place as a function of the relative response strengths, and in so far as *E* can control or measure these strengths, the kind of behavior which will occur could be predicted. In simple motor responses, data are available for such predictions (340).

Gradients of approach and avoidance. The problem of prediction in a spatial type of conflict is further complicated by the fact that there are gradients of approach and gradients of avoidance which do not necessarily have the same slope. These gradients imply that the strength of the response tendency to approach (or to avoid) varies inversely as the distance of the organism from the incentive. This has been shown in animals by measuring the speed of running and the strength of pull when the animal is placed at various distances from a goal (41). The avoidance tendency has been shown to vary in strength as a function of distance from the avoidance stimulus by placing the animal at various distances from this stimulus and measuring the speed and force with which the animal retreats (47).

Miller's analysis of various data (259) on gradients of approach and avoidance also makes it clear that the avoidance gradient is steeper than the approach gradient. The implications of this may be suggested by an illustration. Suppose it was determined that the strength of an approach tendency when an animal was just out of reach of a piece of food was the same as the strength of an avoidance tendency at a point right next to an electric grid which had been used to shock the animal. Now we put the food on the grid and place the rat, let us say, 5 feet away. Will he approach or retreat? If the avoidance gradient has a steeper slope than the approach gradient we would expect the animal to approach to the edge of the grid at which point the tendencies are equal. Thus, on the basis of differences in the slope of approach and avoidance gradients, *relative response strengths will also be determined by distance of the organism from a negative or positive stimulus situation.*

Let us take another illustration to show how neatly the difference in slope of approach and avoidance gradients will allow for prediction of behavior. Miller (259) reports these observations from work on a straight-alley maze. First, Miller determined the strength of pull of a rat *away* from a shock at one end of the maze. Then he determined the strength of pull *toward* food. Both measurements were made in close proximity to the shock source and food respectively, and under these conditions the avoidance tendency was stronger than the approach tendency. Next it was shown that at some distance away from the food and shock source the approach tendency was stronger than the avoidance tendency. This implies that the two gradients cross each other; hence, have different slopes.

We may schematize this situation by assuming the straight-alley maze to be the abscissa of a graph on which the ordinate represents response strength (Fig. 45). After the approach and avoidance tendencies were set up by the rat's receiving food and shock respectively in the left end of the maze, what prediction would we make concerning his behavior if he is then placed in the starting box at the right end of the maze? Initially, the rat should start running toward the other end—the food box end—since the approach tendency is strongest. However, at the point where the two gradients intersect (dotted line), the rat

should halt, because the two tendencies of approaching and avoiding are of equal strength. This is exactly what the animal did in one such situation. He started running down the alley and then literally skidded to a halt, subsequently showing alternate approach and avoidance behavior.

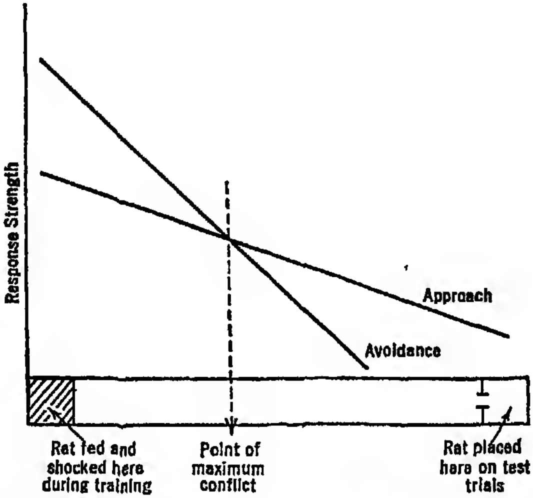


FIG. 45. Prediction of conflict behavior on the basis of differences in slope of approach and avoidance gradients. See text for complete explanation.

Several other predictions have been made and confirmed concerning the implications of the difference in slope of the approach and avoidance gradients. Distance from a goal, or from a negative incentive, appears to be a clear variable in determining response strength.

Known and Probable Variables

When we consider the problem of variables, we have one general question to ask: "What are the factors which cause conflict responses to vary in amount or degree?" Let us set down in step form the relationships between manipulable stimulus variables and conflict responses.

1. *Relative response strengths of incompatible tendencies.* The more nearly equal the opposed tendencies the greater the conflict. *E* may manipulate the strength of a given tendency by:

a. Varying number of training trials.
b. Varying amount of positive incentive, e.g., amount of food, and by varying amount of negative incentive, e.g., amount of shock.
c. Varying approach tendencies in terms of strength of motivation, e.g., hours of deprivation. No equivalent manipulation is possible for avoidance tendencies.
d. Varying distance from negative and positive stimuli. This would be based on empirically determined differences in slope of approach and avoidance gradients for given negative and positive incentives. It is not known, for example, whether or not different kinds of negative incentives produce different slopes of their respective avoidance gradients.

2. *Nature of the incompatibility.* Nearly all studies have tended to show that an approach-approach situation produces very little conflict behavior as compared with the other three types of situations. Among themselves the other three situations have been shown to produce varying amounts of conflict (e.g., Brown's study). We cannot be sure that this is due to the intrinsic nature of the situations or whether other factors, such as relative response strengths, have also varied.

3. *The absolute strength of incompatible tendencies.* We may keep relative strengths of tendencies constant and vary absolute strengths by the methods indicated in 1-*b*, and 1-*c*. Evidence tends to *suggest* that the greater the absolute strengths, the more severe the conflict responses. We may consider as severe conflict responses those which are indicated by seizures, strong fixations, withdrawal to a point of starvation, and strong emotional responses. These will be indicative of severe conflict as compared, for example, with vicarious trial and error. The latter may occur with the former, but when it occurs alone, conflict is small in amount.

4. *Duration of exposure to conflict situation.* Some evidence would support the principle that in situations in which physical restriction is high, the length of exposure is directly related to amount of conflict. Let us quote Liddell on this matter:

> The daily period of self-imposed restraint to which the animal is subjected in the conditioning laboratory, together with its relinquishment of initiative and spontaneity within the testing period, we now believe, lays the foundation for the pathological outcome of long-continued training (219, p. 395).

There is another problem concerning the variable of duration which has received little attention. We refer to the resolution of conflicts. What mechanisms, for example, bring about the resolution of the conflict when the rat is "caught" in an approach-avoidance situation in a straight-alley maze? What happens as a function of duration of conflict, which resolves it? Do the approach and avoidance tendencies extinguish or weaken? Are there other responses which interfere with conflict responses? Very little is known about these problems.

5. *Escape possibilities.* This applies most specifically to the avoidance-avoidance situation. It seems clear that the greater the escape possibilities the less is the conflict as measured by all other indices except withdrawal.

EMOTION AND MOTIVATION

Throughout this and the previous chapter, and to a lesser extent in the chapter on motivation (VI), we have used the term *emotion* without defining it. We stated only that it was used as a synonym for motivation. We are now prepared to consider this matter more fully.

The status of our knowledge concerning emotional processes as usually treated can only be described as lacking or confused. There are physiological changes concomitant with what are commonly called emotional states, but there is little agreement as to the relationship between these physiological changes and a particular emotion (6). For our purpose, then, it will be important to get a point of view or conception of emotions which will integrate with other concepts with which we have dealt. The point of view to be suggested here is for the sole purpose of giving a working orientation for the prediction of behavior. It makes no attempt comprehend the physiological processes involved.

It is proposed that we conceive of emotion as a variable state of the organism, the variation being described in terms of a dimension of emotional tone or tension. We do this without attempting to make a *critical* differentiation between emotional states. This dimension would extend from a point of low emotional tension (sometimes described as apathy, lethargy, ennui,

and so forth) to a point of high emotional tension (sometimes described as excitement, anger, rage, and so forth).

It is further proposed that we conceive of *motivation* variations as lying along exactly the same continuum. This indicates that in terms of functional relationships which hold between variations in motivation and performance, and between variations in emotional tone and performance, we are dealing with the same process, or at least different aspects of the same process. We have used changes in tension as the basic measurement from which we infer changes in motivation. We may think of the same measurement as the basic index for inferring changes in emotional tone, and if we do this, we cannot scientifically distinguish between the two.

If past practices have been interpreted correctly, strong emotions, such as rage, fear, or anxiety, have been used to describe what we have referred to as the high end of the motivational continuum, and such states as sorrow and dejection have been used to describe the low end of the motivational continuum. There seems to be no reason to keep the two conceptual categories separate, especially since the data, although somewhat skimpy, show that in terms of performance changes, both motivation and emotion show the same relationships. Briefly, let us sum up this evidence which has been scattered throughout the last three chapters.

We may state the general principle that when an organism is working at either extreme of the motivation-emotion continuum, a decrement in performance will result as compared with middle-of-the-continuum performance. How much decrement occurs will depend upon many variables, such as type of task, stage of learning, etc., but the *general* principle holds. It is a rare case in which the optimum motivation-emotion condition is at the high end of the continuum. Furthermore, it is doubtful if an organism can operate at the high end for any long period of time without physical breakdown. We might even suspect that the upper quarter of the continuum represents the "ulcer range."

Let us sample the evidence concerning the organism working at the high end of the continuum:

1. Induced tension beyond a certain point produces a decrement in learning. We have conceived of induced tension as a method of increasing

motivation. It has been suggested by some that learning is inhibited at high tension levels because of the emotional states which appear. We are suggesting that these emotional states and high motivation are one and the same thing.

2. Problem-solving ability will decrease under so-called strong emotional responses following frustration. Non-adjustive, bull-dozing behavior emerges when over-stimulation occurs.

3. Strong punishment leads to panic-type responses and it is observed that the variability or plasticity is reduced. This rigidity reduces the probabilities of adjusting to a new situation.

Experimental evidence of the relationships between performance and the low end of the motivation-emotion continuum is consistent throughout. We have seen several instances in which frustration appeared to lead to apathy and dejection, performance scores decreasing concomitantly. This, we suggest, is motivational-emotional deterioration. Anecdotal and case history evidence will further substantiate the general relationships between emotion and motivation and the two extremes of the continuum.

Without ever considering the extremes of the continuum, and how these extremes are correlated with performance decrement, we have evidence that the so-called emotions have the same functional properties as do motives, i.e., they energize, direct, and sustain behavior (cf., Leeper, 212, for a summary of these functional properties as well as an excellent discussion of the whole problem of emotions and their definition). For example, Miller (260) has recently shown that if a rat, shocked in a box for several trials, is allowed to escape to another box after each shock, the avoidance response (fear) to the situation will perpetuate itself so that it may be used to establish a new response of lever-pressing. The latter is true even though the shock as such has long since been removed. The stimulus situation associated with the shock became effective in producing escape motivation, and this motivation served as the basis for the acquisition of a new habit. This, of course, is a clear case of derived motivation called fear. Anxiety, commonly considered an emotional state of undercurrent fear, is quite generally accepted as a fundamental motive in human adult behavior; your own self-analysis will assure you of this.

Are there motives which are not emotional in nature? We do not think so. Our conception is that so-called emotional states represent the high and low end of the motivation-emotion continuum, and that any motive, when it becomes strong, will produce behavior which normally would lead us to infer an emotional state. The important point is that emotions and motivation are functionally equivalent—are interchangeable when we want to predict behavior. We propose to think of them as one and the same thing.

SUMMARY

1. Conflict is a method used to produce frustration. By this method, two incompatible response tendencies are simultaneously elicited.

2. Four general conflict situations were schematized in spatial terms. These were: (*a*) approach-approach, (*b*) approach-avoidance, (*c*) avoidance-avoidance, and (*d*) double-approach-avoidance.

3. General experimental methods were outlined for setting up each of the above kinds of conflict situations.

4. Methods of response measurement and analysis were discussed at some length. As we would suspect, many responses in the conflict situation are the same as those responses to frustration induced by other methods. The conflict situation has additional responses which appear to be intrinsic to it. All response deviations are actually or assumed to be determined by comparison with a control condition in which there is no incompatibility of responses. The following response measures were illustrated: (*a*) latency, (*b*) vicarious trial and error, (*c*) withdrawal, including both psychological and physical withdrawal, (*d*) compromise responses, (*e*) seizures, (*f*) performance changes, including regression, fixation, and discrimination breakdowns, (*g*) general observations.

5. Basic variables were discussed and their relationship with amount of conflict stated. These variables are: (*a*) relative response strengths, (*b*) nature of incompatibility, (*c*) absolute strength of responses, (*d*) duration of exposure to conflict, and (*e*) escape possibilities.

6. Finally, a proposal was advanced that we view emotions as continuous with motives, and speak of a motivation-emotion continuum.

We have seen that conflict studies have similarities to the psychophysical studies considered in the early chapters. We shall see in the next chapter on transfer of training that de-traumatized conflict is common. In all studies of learning where interference among responses is observed, we are probably dealing with conflict of a less traumatic nature than discussed in this chapter. Basic principles for the prediction of behavior should be common to all responses where interference is noted.

CHAPTER IX

Transfer of Training

INTRODUCTION

A trans-oceanic Lockheed Constellation landed at Shannon Airfield in Eire and started to taxi to the unloading zone. The pilot ordered "flaps-up" and the co-pilot automatically reached for a lever on the left side of the seat. The lever was pulled, the *wheels* came up and the plane dropped 7 feet to the runway and was badly damaged. Investigation showed that the co-pilot had been trained earlier on another type of transport plane in which the flaps-up lever was in the same position as the wheels-up lever on the Constellation. This earlier learned response, pulling lever at left to stimulus "flaps-up," had momentarily displaced the newer response of pulling the proper lever to the stimulus. This is an illustration of transfer of training.

Definition. The influence of previous experience on current performance defines transfer of training. To be most meaningful, however, the definition should be thought of in terms of the operations by which transfer is measured. The experimental paradigm must include two activities or tasks, and an experimental and a control condition. Let X stand for the first task and Y for the second; the operational definition of transfer is then given by the difference in performance on Task Y under the following two conditions:

Experimental Condition:	Task X	Task Y
Control Condition:	....	Task Y

If the effect of Task X is to inhibit performance on Task Y, *negative transfer* has taken place. If the effect of Task X is to facilitate performance on Task Y, *positive transfer* has taken place.

The performance of the confused co-pilot in the above story illustrates negative transfer, as does the evidence for interference between habits given in the previous chapter on conflict. Examples of positive transfer are so commonplace that we often overlook them. We are able to learn algebra more easily as a consequence of mastering arithmetic; we may understand the game of softball more readily because of a knowledge of baseball; we hope we shall be able to grasp the material in this chapter more quickly as a result of having studied the earlier chapters.

As indicated, transfer effects may be positive or negative. We should add, however, that in some situations measurements may show neither positive nor negative transfer, hence we observe zero transfer. Zero transfer may result in two ways: (1) conditions for transfer of any kind are absent, or (2) conditions for both positive and negative transfer are present simultaneously and the two opposed effects cancel. If we consider the entire developmental history of the organism, especially the early years, we can see that the net effects of transfer must be positive or the organism will not survive.

We may think of empirically derived transfer effects as varying along a continuum one end of which is labeled "high positive transfer," the other "high negative transfer," with a point of "zero transfer" in the center. At the empirical level the study of transfer of training resolves itself into a search for stimulus dimensions producing the wide range of transfer effects.

Orientation to topic. We have said that when an organism is placed in a problem situation trial behavior results. This was suggested by the several arrows in Fig. 3, Chapter I. Basically, the study of transfer of training is an attempt to discover the conditions which produce the specific trial responses, i.e., the conditions which cause the appearance of one response instead of another. It is a well-founded belief that when an organism is placed in a relatively new situation it will respond in a way which has been successful in previous situations similar to the present one. If the transferred responses are adequate or correct, positive transfer occurs; if they are more or less persistently inadequate or incorrect, negative transfer occurs. We have studied motivation and the effects of blocking motivated behavior; now, we are ready to study the factors which determine the adequacy or inadequacy of

previous experience as the motivated organism is faced with a problem situation.

The topic of transfer of training has a long and somewhat stormy history. It has always been a central problem in theories of education, since formal education is given under the assumption that such training will transfer in part to real life situations. For decades almost unlimited positive transfer of training was basic educational philosophy. For example, Latin was learned so that English and other languages could be mastered more easily; Homer was memorized so that the faculty of memory could be improved; geometry was a required subject because the study of it would improve reasoning. Shortly after 1900, experimental attacks were launched to test the validity of these assumptions. From many researches we now know that the gross assumptions indicated above are unsound, and that positive transfer in such instances is very limited indeed (198, 233, 422).

Of course, we must still make assumptions concerning the transfer effects from the school room to other situations. These assumptions, however, are much less sweeping in character than was formerly true, since they usually refer to specific skills. This is implied by Stroud when he says: "The only object in teaching a child to count ducks, is to teach him to count; to count anything that can be counted" (371, p. 366). We do not expect training in counting to influence appreciably any skill but counting.

Order of topics. Because of the breadth of this subject, we shall omit from the present chapter certain details of experimental design involving procedures with which we are not familiar. The following chapter, however, will be devoted entirely to a consideration of design problems. The nature of these considerations is such that we must know certain facts about transfer of training before they will be meaningful.

We have suggested that an organism responds to relatively new situations in a manner appropriate to earlier similar situations. Thus, similarity between the first and second tasks becomes a basic dimension causing transfer effects to vary. Much of the chapter will be concerned with ways of manipulating similarity between two tasks. Other variables are actually secondary in determining transfer effects. This is apparent in the following order of topics of the chapter:

1. Methods of response measurement
2. Similarity and transfer: a working orientation
3. Dimensional variations in similarity
4. Non-dimensional variations in similarity
5. The influence of secondary variables

METHODS OF RESPONSE MEASUREMENT

Preliminary design considerations. The paradigm which operationally defines transfer also describes a basic experimental design. The effects of transfer are measured in terms of differences in mean performance on the experimental and control conditions on Task *Y*. Design Method II (Chapter V) requires that the groups be matched, one group serving in the control condition and the other(s) in the experimental condition. Matching may be on the basis of a highly correlated task or on the initial trials of the task to be used in the experiment. The latter is accomplished by giving both groups a few trials or a short period of work on Task *Y* at the outset. Then, the experimental group is given work on Task *X*, and finally, both groups are measured again on Task *Y*. Diagrammatically, this would appear as follows:

Experimental: Task *Y* .... Task *X* .... Task *Y*
Control: Task *Y*.............................. Task *Y*

Differences in performance of the two groups on the final test of Task *Y* would be a function of Task *X*. It should be recognized that the use of the foretest on *Y* to obtain data for matching will, to a certain extent, obscure the total transfer effect of *X* on *Y*. This is true because both groups will have had some acquaintance with Task *Y* before the transfer test is made. Moreover, initial experience with Task *Y* will probably have some transfer effect on Task *X* in the experimental condition, and this may not be desirable in some experiments.

Under certain conditions it is possible to measure transfer effects by the use of a single group of *S*s serving in only one condition; namely, the experimental condition as outlined in the experimental paradigm given to define transfer. The necessary restriction is that the two tasks, *X* and *Y*, be of equal difficulty. If they are of equal difficulty a comparison of the scores on Task *X* with those on Task *Y* will indicate the amount of transfer. We seldom have

two tasks which are known to be of equal difficulty, so that in actual practice the method is infrequently used.

Units of measurement. For most tasks, one of the following four units is applicable: (1) trials to attain a given level of mastery; (2) time to attain a given level of mastery; (3) level attained after a given time or given number of trials; (4) number of errors made in attaining a given criterion of mastery, or number of errors made during a given period of time or for a given number of trials. More than one measure is often employed in a single experiment. The amount and direction of transfer (positive or negative) will be determined by comparing the mean performance on the control condition with the mean performance on the experimental condition on the common task.

The raw score units are sometimes expressed in percentage figures. Such a procedure is necessary, of course, if the transfer effects of one set of tasks are to be compared with those of another set in which the raw score units are different. There are several ways of expressing the transfer in percentage figures (102), but one illustration will do for our purpose. Let C stand for the mean of the control group on Task Y, and E the mean of the experimental group on the same task. The amount of transfer expressed in per cent would be:

$$\frac{C - E}{C} \times 100.$$

If the raw score unit of measurement is such that the larger the value the poorer the performance (as in trials to reach a given criterion of mastery), the resulting per cent value will be negative if the transfer is negative and positive if the transfer is positive. If the unit of measurement is such that the larger the value the better the performance (as in level of mastery for constant number of trials), the signs will be reversed.

In a given transfer situation there may be conditions producing positive transfer and other conditions producing negative transfer. Furthermore, the influence of these conditions may vary markedly, depending upon the level of performance on the second task, Task Y. The measurement we take at any point in performance on Task Y will nevertheless show only the *net* effects of the two influences; the measurement will show *either* positive *or*

negative transfer, but not both. Consequently, to evaluate the interaction of positive and negative transfer *we must take a series of performance measures on Task Y.*

SIMILARITY AND TRANSFER: A WORKING ORIENTATION

Similarity is a multiple variable in that it has several independent operational definitions. Furthermore there are (usually) several components of two tasks between which a given similarity relationship may hold. Thus, the problem of relating transfer and similarity is a complex one.

The attempt to simplify the study of similarity and transfer involves two aims: (1) making the exemplary transfer situation as analytical as possible; and (2) providing a theoretical orientation which will unify the results from various similarity dimensions. The first objective may be accomplished by use of the paired-associate learning situation, and the second objective may be realized by an application of the principle of generalization.

The Paired-Associate Technique

The paired-associate technique, used primarily with verbal material, is a method wherein *S* is presented with a series of pairs of words, the first member of the pair being the stimulus and the second member the response. *S* is to associate each response with its stimulus, so that when he sees only the stimulus he can make the response. Thus, in the instance, *DESK—ARMY*, he would learn to say *ARMY* when *DESK* is presented.

Paired associates are customarily presented on a *memory drum*, a device which can expose one pair of words at a time, and each successive pair for the same time interval. Typically, the stimulus word of a pair is presented for 4 seconds. During the first 2 seconds of this period the stimulus appears alone, but during the last 2 seconds the response is shown with the stimulus. Then, the stimulus word of the next pair is shown, and so on down through the list. *S* is to learn to call out the appropriate response to each stimulus word before the response word appears. He has 2 seconds (when the stimulus is presented alone) in which to anticipate the response. These time intervals are arbitrary, of course, but are the ones which have become fairly standard.

The number of pairs in a paired-associate list has varied, but from 8 to 15 have commonly been employed. Before the first presentation of any list *S* cannot know what the correct responses are. So, he merely studies the pairs on the first trial and tries to associate each response with its stimulus. On the second trial he may remember some of the responses and will call out the responses during the 2-second anticipation period. *E* records correct responses as well as incorrect ones. *S* goes through the list time after time until he has reached the criterion of performance which *E* has established. The words are presented *in a different order* on each trial so that *S* cannot learn which stimulus is coming into the aperture of the drum before it actually appears, and so that he will not learn the response words serially by ignoring the stimuli.

In studying transfer we must have at least two tasks; in paired-associate learning these would be two lists of word pairs. While we shall orient our thinking around this technique and its implications, we shall by no means restrict our illustrations to studies in which it has been used.

Generalization and Transfer

In Chapter VIII we saw that generalization could be used to explain a number of conflict phenomena. Whenever changes in performance are measured as a function of changes in similarity of conditions, generalization is a useful explanatory concept. Because of the several ways in which similarity may be varied, it will probably help unify and simplify our thinking if we assume that all these variations are accompanied by changes in degree of generalization.

Stimulus generalization. When stimulus *A* gains power to elicit response *B*, other stimuli similar to *A* may also be shown to have some tendency to elicit *B*. This is stimulus generalization. Let us call these similar stimuli *A1*, *A2*, and *A3*, in which *A1* is most similar to *A*, *A2* next most similar, and *A3*, least similar. The generalization gradient which we have preveiously discussed implies that *A1* has the greatest probability of eliciting *B*, *A2* the next greatest, and so on.

Response generalization. When stimulus *A* gains power to elicit response *B*, it will also tend to elicit responses which are in some

way similar to B. This is called response generalization. Responses which are similar to B may be called $B1$, $B2$, and $B3$, representing decreasing similarity with B.

There is little systematic experimental evidence concerning response generalization (153). We can see that a gradient of response generalization could be derived empirically only after studying responses which had been made to a given training stimulus. For example, we would use stimulus A as the training stimulus for eliciting response B. Then we would present stimulus A as a test, keeping a record of all responses given to it. A great many *S*s (or responses) would be required. Finally, we would scale the responses as to similarity with response B. If the frequency of responses $B1$, $B2$, and $B3$ were directly related to their similarity to B, we would have independent evidence (i.e., from the scaling) for a gradient of response generalization. Such an experiment has not been performed, so we can only suggest the probable existence of such a gradient.

An illustration of response generalization is provided by a study on retention (393). *S*s learned two lists of paired associates. Each member of the pair of each list was a two-syllable adjective. A pair in the first list may be noted *A-B* (response B to stimulus A) and the comparable pair in the second list may be noted as *A-C* (the same stimulus as used in the first list but a completely different response). Sometime after learning the two lists, S was asked to give the first of the two responses which occurred to him when the stimulus A was presented—a word association test. Under such conditions the B or C response occurred most frequently, but in an appreciable number of cases responses occurred which were similar to C or B. Thus, the response *resting* occurred whereas *restful* was a B or C response; *misty* appeared in place of *filmy; inane* instead of *witless; gaiety* instead of *festive.* The most reasonable interpretation of these findings is that these responses occurred as a consequence of being similar to the learned response; thus, they demonstrate response generalization.

Fully recognizing that we do not have systematic evidence of a gradient of response generalization which has been independently derived as in the case of gradients of stimulus generalization, let us assume that such a gradient of response generalization exists. Then we may see how these two generalization gradients will aid

us in understanding the results of experiments in which similarity has been varied.

Stimulus generalization and transfer. *S* learns two lists of paired associates in which the responses are identical. We shall symbolize these identical responses by *B*. The stimuli of the two lists are completely different, so we shall call them *A* and *C*. Our first list of items would be *A-B*, the second *C-B*. Now, in a series of experimental conditions, the stimuli of the second list are made

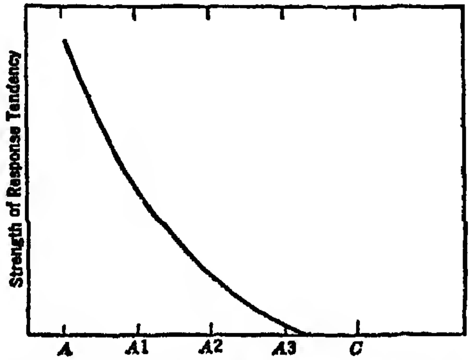


FIG. 46. Gradient of stimulus generalization as it is assumed to exist following the learning of the first list, *A-B*. *A1* to *C* represent stimuli of decreasing similarity to *A*. The ordinate represents the strength of response tendency of these stimuli to elicit *B* in learning the second list.

more and more similar to the stimuli in the first list. Thus, we might have four conditions in which *A-B* is always learned as the first list, then *C-B* as the second list in one condition, *A3-B* the second list in another, *A2-B* in another, and *A1-B* in the fourth. All stimuli in the second list would be changed in the same fashion. What would we predict concerning the transfer effect in learning the second list under the various conditions? To help us make this prediction let us visualize a gradient of stimulus generalization as in Fig. 46. Since *A* is the stimulus in the first list, our other stimuli must fall at differing distances from it along the similarity dimension.

Note that in the verbal situation we deal with only one "wing" of the gradient. We have drawn this gradient so that it is concave in shape instead of a straight-line as in Chapter VIII. The data we will consider later suggest that this is the most probable shape

for verbal materials. The gradient implies that not only has A gained power to elicit B; but that $A1$, $A2$, and $A3$ have also gained some response strength toward eliciting B as a result of stimulus generalization from A. Stimulus C, being outside the gradient, has gained no tendency to elicit B as a consequence of the learning of A-B. These are the situations as they exist at the *end* of learning the A-B list, *before S starts to learn any one of the second lists.*

With the gradient of stimulus generalization well in mind, our prediction concerning the relative speed of learning the second list in the four conditions become obvious. When the second list is C-B, no positive transfer should occur as a consequence of stimulus generalization, although some positive transfer may occur as a consequence of other factors. In learning $A3$-B, some positive transfer should occur since $A3$ as a stimulus already has a small amount of associative connection with B as a consequence of learning the first or A-B list. When the second list is $A2$-B, learning should take place at a still faster rate, and $A1$-B should be learned most rapidly of all. In summary of our considerations, we may state the general principle as follows: *When old responses are learned to new stimuli, the more similar the stimuli the greater the positive transfer.* We should never get negative transfer in such a situation, although with well-practiced *S*s we might get very nearly zero transfer when we learn C-B following the learning of A-B.

It is upon the above principle that our belief in positive transfer from the school situation to real life situations is based. The child who can count pictures of ducks will probably do fairly well in counting real ducks on the pond.

Response generalization and transfer. To conceptualize response generalization as it relates to transfer in learning paired-associate lists, let us consider the situation in which the stimuli of the two lists are identical, and the responses vary from complete dissimilarity to high similarity. The first list will always be A-B, and for four conditions of similarity of the first and second list responses we have A-C, A-$B3$, A-$B2$, and A-$B1$ from low to high similarity respectively.

Figure 47 shows the gradient of response generalization as it is assumed to exist after learning the first or A-B list. The figure

shows that the strongest response tendency to *A* is *B*, while *C* has no associative strength to *A*. Now, what will we predict with regard to the transfer effect under the four conditions of similarity?

When the second list is *A-C*, the tendency of stimulus *A* to elicit response *B* is very strong, ***but actually is an erroneous response tendency in learning the second list.*** Consequently, this situation should produce negative transfer, since the old response tendency (*B* to *A*)

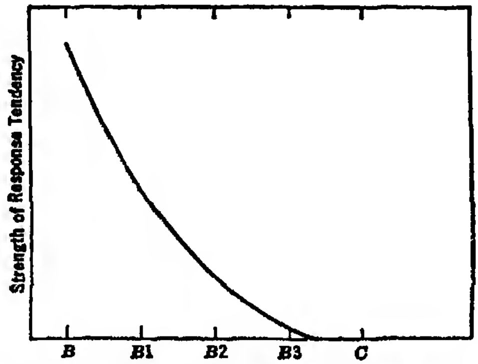


FIG. 47. Gradient of response generalization as it is assumed to exist following the learning of the first list, *A-B*. *B1* to *C* represent responses in the second list of decreasing similarity to *B*. The ordinate represents the strength of response tendency of stimulus *A* to elicit these responses in the learning of the second list.

will interfere with the learning of *C* to *A*. The data of many experiments strongly support this inference. In fact, we may say that this paradigm, learning a new response to an old stimulus, is the ***optimum condition for producing negative transfer.*** As we go from the first list, *A-B*, to the second list, *A-B3*, the strongest tendency should still be that of calling out response *B*. As a result of response generalization, however, *B3* has some associative connection with stimulus *A*. Consequently, the negative transfer should be less than when the second list is *A-C*. Similarly, as the response becomes still more similar to *B*, i.e., *B2* and *B1* the less should be the negative transfer in learning the second list. Indeed, other factors which produce positive transfer may be strong enough so that we will be unable to measure negative transfer when the second list responses are quite similar to the

first list responses. Complete positive transfer will occur when the second list is *A-B*—the same as the first list.

It is apparent that negative transfer is produced by interference of first list responses with attempts to learn second list responses. Therefore, if learning of the second list is to take place at all, the responses from the first list must be inhibited in some fashion. This inhibition is believed to take place by *differential reinforcement*. We mean by this that during the learning of the second list the wrong responses from the first list are no longer rewarded (symbolically), while the new responses of the second list are rewarded if correct. The result should be that the second list responses become stronger than the interfering first list responses. Furthermore, although the data are fragmentary on the matter, it is possible that there may be an absolute weakening of the first list responses as a consequence of their non-reinforcement during the learning of the second list. For example, if an association in the first list has four units of strength at the end of learning, it may be weakened as a result of non-reinforcement during the learning of the second list so that its associative strength is reduced to two units.

The fact that negative transfer is produced by the interference of old responses with learning of the new suggests that the old responses may show themselves during the learning of the second list. Several studies show that this is actually what happens, although the frequency of these erroneous responses is not great. Much of the interference seems to be implicit, i.e., *S* recognizes the wrong response tendency but cannot think of the correct response. When the errors are explicit, that is, when first list responses actually appear during the learning of the second list, they are called *intrusions*.

Other predictions. The two above situations represent what might be called *basic transfer paradigms*. There are, however, other ways in which similarity may be varied between components of the two paired-associate tasks, and results of some of these variations may be predicted by studying generalization gradients.

Let us assume that in one condition *S*s learn *A-B*; *A-C*. As discussed above, we know that negative transfer would be exhibited in learning *A-C*. Now, on successive conditions we keep the responses (*B* and *C*) completely different and decrease the

similarity of the stimuli (*A1*, *A2*, *A3*, and *D*). The more dissimilar the stimuli in the two lists become, the less and less will the *A-B* (first list) connections interfere with the learning of the second list. Consequently, when the stimuli of the two lists become completely different so that *S* learns *A-B*, *D-C*, we would predict zero transfer effects as a result of our given dimension of similarity.

What predictions would you make if one condition is *A-B*, *C-B*, and on successive conditions the stimuli remain dissimilar while the similarity of the responses is decreased?

We have seen how generalization may be used to predict the outcome of transfer experiments where similarity is systematically varied. Now, let us turn to experimental studies in order to determine ways in which similarity is defined and varied.

DIMENSIONAL VARIATIONS IN SIMILARITY

Variation in Formal Identity

Formal identity in paired nonsense syllables. Nonsense syllables are combinations of three letters which do not make a meaningful word. Usually, these syllables consist of two consonants and a vowel, with the vowel in the middle, although syllables consisting entirely of consonants have been used. Illustrations of nonsense syllables are: *JOH; XAN; VAW; KER; QUC.* Illustrations of consonant syllables are *PHM; FRK; SCM; JNX.* In learning nonsense syllables *S* is sometimes instructed to spell the response rather than pronounce it since some syllables almost defy pronunciation. The words are customarily presented on a memory drum in the manner already described.

To vary the degree of formal identity between the components of two nonsense syllable lists, *E* varies the number of identical letters. We may illustrate this with a study by Bruce (44). He used many conditions, but four of these will illustrate the method of varying similarity.

	First List	Second List
Stimuli Identical, Responses Different:	*DOW–NAZ*	*DOW–SIV*
Stimuli Identical, Responses Highly Similar:	*DOW–NAZ*	*DOW–NAM*
Stimuli Different, Responses Identical:	*DOW–NAZ*	*SIV –NAZ*
Stimuli Highly Similar, Responses Identical:	*DOW–NAZ*	*DOR –NAZ*

You will note that in the *highly similar* stiuations two letters of the responses, or of the stimuli, are identical. There might also be

situations in which only one letter is repeated, thus forming an intermediate step between dissimilarity and two degrees of identity. It is not implied that variation of number of identical components produces a uniform change in amount of transfer.

The transfer produced by Bruce's conditions was very much like that which we would predict from our analysis of generalization gradients. The first situation, learning different responses to identical stimuli, produced clear-cut negative transfer. The second situation, learning highly similar responses to identical stimuli, resulted in negligible transfer. The third situation, in which responses were identical and stimuli different, gave slight positive transfer, and the fourth situation produced high positive transfer. We should again point out that there are other variables which influence the amount and direction of transfer. These other factors, which we are to consider later, do not appear to influence greatly the *relative* values for transfer as a function of similarity, but may change the absolute values considerably.

Formal identity in serial learning. Paired-associate learning may be likened to the learning of a foreign language vocabulary. Each foreign word (the stimulus) is to be followed by its English translation (the response). Serial learning, on the other hand, may be likened to the learning of the alphabet, the number series, or successive bus stops. In serial learning the items appear always in the same series, *A*, *B*, *C*, *D*, and so forth. *S*, to learn the list, must learn to anticipate (call out) *B* when *A* comes into the aperture of the memory drum; he must learn to say *C* when *B* comes into the aperture, and so forth. The list is repeated until *S* reaches a chosen criterion of mastery. It is common in both paired-associate and serial learning to allow *S* a short rest (6 to 10 seconds) between each successive presentation of the list.

The similarity between serial lists is much less readily analyzed than it is between paired-associate lists. Each item in a serial list is both a stimulus and a response (except the last item which is only a response, and the first item which is only a stimulus). In serial learning there is no distinct stimulus item in the first list which may be compared with a distinct stimulus item in the second list. The same is true concerning responses.

In spite of this, we can show, in general, that serial lists can be varied in terms of amount of formal identity, and that this is

related to amount of transfer. An illustration is given in a study by Melton and von Lackum (257). Two basic conditions were employed, one in which the second list was highly similar to the first, and the other in which the second list was quite dissimilar to the first. High similarity was achieved by constructing the two lists with the same nine consonants. Each list was 10 syllables in length, making a total of 30 letters. This means that on the average each letter was repeated 3.33 times in each list. Dissimilarity of the two lists was produced by making the second list with nine completely different consonants. Illustrations of lists used in the two conditions are as follows:

First List	*Second List: Highly Similar Condition*	*Second List: Highly Dissimilar Condition*
XCM	FTM	JSN
KTQ	HCR	GBD
HFX	XQT	LNJ
CKM	KRM	ZPB
XRF	CTF	SDG
QHC	QXK	LJP
KXT	MHT	ZGN
MQF	RKF	BSJ
CTK	CMQ	PZL
RMH	TXH	NDB

In the first list only the letters *C*, *F*, *H*, *K*, *M*, *Q*, *R*, *T*, and *X* were used in making the 10 syllables; the same letters were used in constructing the highly similar list. We might expect that the associations formed during the learning of the first list would interfere with the formation of new associations in learning the second. The results show that this is probably what happened, since significant negative transfer was measured in the learning of the second list after the first list had been presented for 5 trials. When the second list was highly dissimilar (having no letters in common with the first), no evidence of either positive or negative transfer was found. Thus, it appears that formal identity of letters between two lists is one way in which amount of transfer can be varied. Within limits we would presume that the greater the degree of formal identity, the greater would be the negative transfer.

Another source of interference between serial lists should be mentioned. It has been noted that a word in the first list is most

likely to interfere with a word in the second list when both occur at the same *serial position*. Thus, if *PBN* occurs at the fourth position in the first list it is more likely to interfere with the fourth syllable in the second list than with syllables in other serial positions, assuming similarity factors are constant. The fact that *PBN* may sometimes be given at the fourth serial position while *S* is learning the second list is indicative of serial position identity as a source of interference. It is very difficult, however, to evaluate the extent of the influence of such interference. Results in the above study in which the learning of dissimilar lists showed neither positive nor negative transfer suggest that if serial position identity produces negative transfer, it is not great enough in amount to overcome factors making for positive transfer.

Formal identity in motor interference. Some of us have gone through school without fully avoiding the hideous sin of splitting infinitives so that when it is mandatory to répair these consistently, considerable interference is evident. If we carry our cigarettes in one pocket for a considerable period of time and then shift them to another, many false moves will be made to the stimulus, "want a cigarette." The interference which occurs in the case of the infinitives may be thought of as relatively non-motor when compared with the interference which occurs in reaching for cigarettes. At this point we wish to consider simple motor interference as a function of formal identity of movements.

Card-sorting is a task commonly used to demonstrate motor interference in the laboratory. *S*s learn first to sort a deck of ordinary playing cards into four boxes labeled respectively, hearts, diamonds, clubs, spades. Then *E* changes the position of the boxes and *S* again sorts the deck. *S* makes many false moves in the first few trials after the change in position of the suit boxes. He must learn, for example, that the heart box is now where the spade box used to be. We would suspect that the amount of negative transfer would be directly related to the number of changes in movement required to sort the deck. The expected variation in amount of transfer is demonstrated nicely in a study by Crafts (72).

Crafts used Flinch cards, on which the numbers occur in bold type. Nine numbers were used, and each number appeared eight times in the deck of 72 cards. A system of nine boxes, each $4^1/_2$

inches square, was constructed in tier fashion. *S*'s task was to sort the cards as rapidly as possible into the appropriate boxes, each of which was numbered to indicate the cards to be placed in it. The response measure was the time required to sort the deck, one sorting constituting a trial.

After brief preliminary practice, all *S*s were given 8 trials. Four groups of 16 *S*s each were matched on the basis of scores made on the 8 trials (Design Method II). After the matching, every group was given two additional trials under the following conditions:

Group I:	Position of *no* numbers changed.
Group II:	Position of 3 numbers changed.
Group III:	Position of 6 numbers changed.
Group IV:	Position of all 9 numbers changed.

Note that as we go from Group I through Group IV there is an increase in the number of new responses which *S* must make to the old stimuli (the cards). Group I, in which no numbers were changed, served as a control.

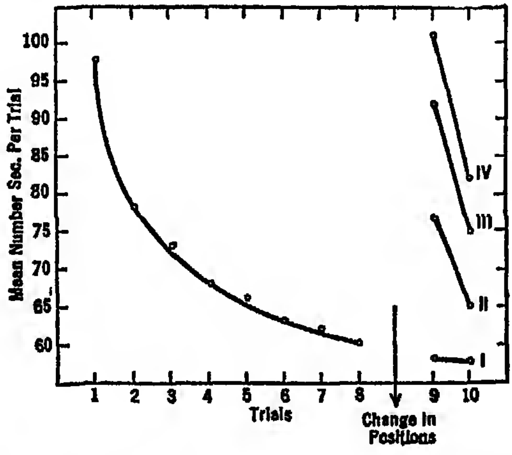


FIG. 48. Transfer in card-sorting as a function of formal identity. See text for explanation of groups. Data from Crafts (72).

The results of Crafts' experiment are shown in Fig. 48. The four groups were well matched, so that one curve shows the performance curve for the first 8 trials for all groups combined.

When there was no change (Group I) there is relatively little change in the performance curve. When the position of all numbers was changed, a great deal of negative transfer occurs. The performance of Groups II and III is intermediate between the two extremes. These results show that as the number of old habits to be broken or inhibited increases, the amount of negative transfer increases.

Actually, the negative transfer spoken of in Crafts' experiment is relative negative transfer. Only Group IV did poorer after the change of positions than they had done at the start of the experiment. All other groups show absolute positive transfer when compared with their performances at the start of the experiment. That is, all groups except IV showed better performance after the change than they did on the first of the preliminary 8 trials.

We have considered three illustrations by which formal identity of two tasks may be varied experimentally. We have seen that formal identity is manipulated by varying the number of duplicated units in the first and second tasks. This of necessity must take place by discrete steps. Subsequent methods of varying similarity will show variation along a continuous dimension.

Formal Similarity of Figural Designs

E may construct designs which vary in amount of formal similarity to a standard design. Usually these designs are relatively meaningless in that they do not immediately suggest a known form or object. Such forms have been used systematically to study transfer as a function of stimulus similarity in paired-associate learning. As our illustration of this method we will use the work of Gibson (107) and Hamilton (138). Both *Es* used exactly the same materials.

Gibson drew 12 forms to be used as standard stimuli. Then a large group of additional forms were made up which resembled the standard forms in varying degrees. Preliminary to the main experiment a group of judges was asked to sort this large group of forms into 12 piles, each pile being represented by a standard form. The judges were to put a form in a pile if it resembled the standard of that pile more than it resembled one of the other standards. Finally, the judges were instructed to rank order the forms within a given pile on the basis of degree of similarity to the standard.

The final set of materials consisted of the 12 standards, with three other forms of varying degrees of similarity to each standard. An illustration of one standard form and its three similar forms is shown in Fig. 49.

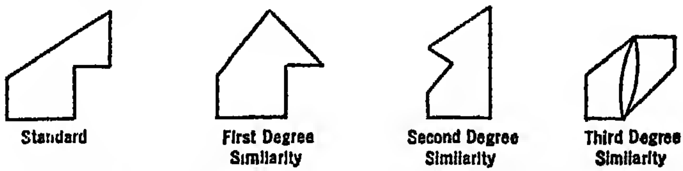


FIG. 49. Illustrations of designs of varying degrees of similarity with a standard. Adapted from Gibson (107).

In Gibson's experiment *Ss* had to associate a nonsense syllable with each of the 12 standard stimulus forms of the first list, and then learn to associate a *different* nonsense syllable when learning the second list. Thus, while the stimuli between the two lists varied as to similarity, the responses were always different (no similarity). In Hamilton's experiment the stimulus items (forms) were the same as in Gibson's, but both lists had the same responses. *S* thus learned to associate the *same* response with stimuli of varying degrees of similarity.

In conducting the transfer experiments the 12 standard forms were used as stimuli in the first list, and in four other conditions the degree of similarity of the second task stimulus with the standard was varied. Four conditions resulted because both Gibson and Hamilton used one condition in which the standard forms were used as stimuli in both the first and second lists.

Let us summarize the four conditions in each experiment:

	GIBSON		HAMILTON	
	First List	*Second List*	*First List*	*Second List*
Condition I	A-B	A-C	A-B	A-B
Condition II	A-B	A1-C	A-B	A1-B
Condition III	A-B	A2-C	A-B	A2-B
Condition IV	A-B	A3-C	A-B	A3-B

The first list, the list with the standard stimulus forms, was presented until *S* correctly anticipated 8 or more of the responses; then the second list was presented until *S* had attained the same criterion. A different group of *Ss* was used for each condition,

and within each experiment the groups were matched for learning ability.

TABLE 21

TRANSFER AS A FUNCTION OF FORMAL SIMILARITY OF FIGURAL DESIGNS

Condition	*Mean Number Trials to Learn*			
	GIBSON		HAMILTON	
	First List	Second List	First List	Second List
I	6.86	7.43	6.73	1.27
II	7.00	5.43	7.40	2.08
III	7.07	5.71	7.40	2.35
IV	7.14	3.93	7.44	3.48

Data from Gibson (107) and Hamilton (138).

The results (Table 21) are presented in terms of mean number of trials to reach the criterion. For comparative purposes, the mean number of trials to reach the criterion on the first list is also presented.

From the results of Gibson's experiment it can be seen that the slowest learning of the second list took place when the stimuli were identical and the responses different (*A-B*, *A-C*). The fastest learning of the second list took place when the two stimuli were quite dissimilar (*A-B*, *A3-C*). No differences were produced by Conditions II and III. Nevertheless, the wide differences at the extremes (I and IV) leave no doubt that Gibson's findings further confirm the basic principles we have discussed previously. When the responses for two tasks are different, the less the degree of stimulus generalization, the faster is the learning of the second task.

Hamilton's results show how similarity of stimuli markedly influence learning when the responses are identical in the two lists. On Condition I it will be noted that the stimuli of the two lists are identical, as are also the responses. Thus, the second list is merely a continuation of the first, and, on the average, slightly over one trial was required to reach the criterion already gained on the first list. The amount of positive transfer becomes less and less as the stimuli become less similar. The results of Condition IV of Hamilton's experiment are very much the same as those

of Gibson's for Condition IV. In line with our previous discussion, this would be expected. We have said that transfer due to generalization would be of small consequence when the stimuli are different and responses the same. This was the situation in Hamilton's Condition IV, whereas in Gibson's Condition IV neither the stimuli nor the responses were the same. Both conditions give about the same results.

In Gibson's experiment only Condition I gave evidence for negative transfer. This is shown by comparing the mean number of trials required to learn the first list (6.86) with the mean number for learning the second (7.43), a comparison which can be made if we assume that the two lists are equally difficult. All other conditions show positive transfer. *Ss* in this experiment were not well-practiced, and as we shall see later, poorly practiced *Ss* have high positive transfer due to factors not associated with specific list similarities. This high positive transfer has probably obscured the negative influence of specific list similarities.

Similarity Relationships of Meaningful Material

The manipulation of similarity dimensions of the kind we have discussed above is relatively straightforward. When we attempt to manipulate meaningful material, however, we have a very difficult task. It is difficult in the sense that *E* has trouble in explaining his obtained results as being due to a single stimulus dimension. Meaningful verbal material may vary along several different dimensions, and even if we know the dimensions, it is a problem to hold all constant except one.

It is possible to isolate dimensions when words are taken singly, and this is a necessary first step in experimentation in this area. But, even after this is accomplished by one of the methods discussed in the chapters on discriminal processes, further difficulties arise in putting the words together in a list. *E* must be sure that the different lists he has constructed do not vary along a new dimension as a result of putting the words together. Any characteristic of the lists (other than the one *E* wishes to manipulate) which causes the rate of learning to vary complicates the problem of discerning specific causal relationships. For example, when words are put together in the list, the similarity among the items *within* the list—the intra-list similarity—must be constant, for this

dimension markedly influences the rate of learning. We shall return to these problems in greater detail in a later chapter. At present it is enough for us to know where the source of the difficulties lies. These problems, while not absent when E manipulates similarity of the kinds we have previously discussed (e.g., formal similarity), are minimal in importance since they can be handled adequately.

There are at least three kinds of similarity involved in meaningful verbal material. These are the similarity of (1) synonyms, i.e., words with similar meanings, such as *horrid* and *ghastly;* (2) homonyms, i.e., words with the same sounds but different meanings—a kind of formal similarity as exemplified by the words *bear* and *bare;* (3) antonyms, i.e., words with opposite meanings, such as *strong* and *weak*. Although the scaling of words has been done only for synonymity, it is conceivable that homonymity and antonymity could be dimensionalized.

Haagen (132, 133) has scaled adjectives to derive several degrees of synonymity and has shown that the amount of transfer is related to this dimension in a fashion that approximates closely the theoretical analysis made earlier in the chapter. He has found that both stimulus and response generalization are important in determining the amount and extent of transfer.

Cofer and Foley (62) have analyzed possible generalization gradients for all three kinds of similarity. They show how it is possible for mediating items to produce generalization from a word to its synonym, and from the synonym to its homonym. That is, the word *plant* is similar to *sōw*, but *sōw* is also a homonym of *so; sōw* thus becomes the mediating item between *plant* and *so*. It is quite conceivable that very complex interactions of generalization gradients may take place in verbal material. The empirical studies of these writers (63, 98) have shown that transfer effects are related to antonymity, synonymity, and homonymity, but the material has not been dimensionalized.

We should remember two points with regard to meaningful material and similarity: (1) transfer is related to the three kinds of similarity, and we will assume the exact relationship to be the same as for formal similarity as previously discussed, and (2) experimentation in this area is difficult because dimensions of the stimulus situation are hard to isolate.

A Summary Graph

From the various data which are available, and from theoretical expectations based on generalization where data are not available, the relationships between amount and direction of transfer and similarity may be summarized as in Fig. 50. Figure 50 is based on paired-associate learning, but may be extended to other situations

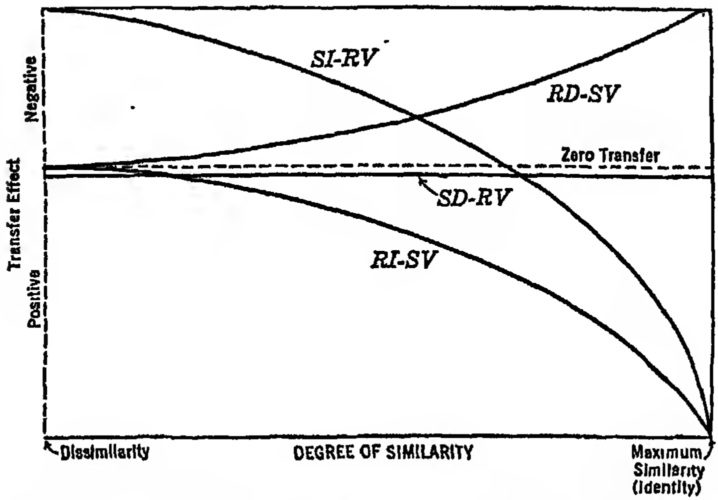


FIG. 50. Relative transfer effects as a function of variation in similarity.
SI-RV: Stimuli identical, responses varied
RI-SV: Responses identical, stimuli varied
RD-SV: Responses different, stimuli varied
SD-RV: Stimuli different, responses varied

when an analysis of similarity relationships between tasks is relatively clear-cut. The graph indicates four ways in which similarity may be varied. One may:

1. Hold the stimuli identical in the two tasks and vary the degree of similarity of the responses (*SI-RV*).

2. Hold the responses identical and vary the degree of similarity of the stimuli (*RI-SV*).

3. Hold the responses completely different (dissimilar) and vary the similarity of the stimuli (*RD-SV*).

4. Hold the stimuli completely different and vary the similarity of the responses (*SD-RV*).

Let us take one illustration of how to read the graph. The curve labelled *RI-SV* indicates that in a series of conditions the responses of the two tasks are identical, but the similarity of the stimuli is varied from low (dissimilarity) to maximum similarity (identity). When the stimuli are dissimilar, no significant transfer effect occurs, but as the degree of similarity of the stimuli increases, positive transfer increases until at identity, 100 per cent transfer takes place.

The values along the ordinate are indeterminant for a general graph, since the curves are intended to show the transfer effects due to similarity only. However, other variables probably will not cause the relative relationships of the curves to change. Roughly, the graph is intended to depict the situation in which a moderate degree of first list learning is used, when *S*s are well-practiced, and when the similarity among the stimuli and among the responses *within* a list is relatively low.

With high degrees of learning on the first task, or with poorly practiced *S*s, very little negative transfer is likely to occur; hence, the dotted line indicating zero transfer would be moved to a point near the top of the graph. Reasons for these qualifications will become evident in later sections of the chapter.

As the graph is drawn, all situations show relationships with similarity except the *SD-RV* paradigm. This is the case in which the stimuli are different and the responses are varied. On the basis of evidence available, and on the basis of generalization, we would expect that variations in the response will produce no appreciable change in transfer effect. The curve has been drawn to indicate slight positive transfer but it might well fall on the dotted line indicating zero transfer. As nearly as can be determined, holding the stimuli different and varying the similarity of the responses produces no change in transfer effects.

NON-DIMENSIONALIZED VARIATION IN SIMILARITY

In this section we shall touch on a variety of transfer phenomena which are a function of similarity between conditions or tasks. In these conditions or tasks, however, the similarity relationships cannot be or have not been analyzed in the manner of the studies reported in the previous section. We shall give at

least one experimental illustration of each phenomenon so that it will be clearly defined in terms of procedures by which it is measured.

Learning-How-to-Learn Transfer

A very common observation is that *Ss*, when learning successive samples of the same kind of material, show marked improvement in their learning ability. Thus, if *S* serves in an experiment in which he learns list of nonsense syllables, one list a day, it is noted that his performance improves from day to day. Descriptively, this phenomenon is called *learning-how-to-learn* (233), and is exactly the same phenomenon to which we have referred earlier as practice effects. In almost any kind of activity in which some adjustment to the situation is required, we will find a practice effect.

The similarity involved in learning-how-to-learn transfer is primarily in material which is of the same general nature or of the same class. The similarity between one task and another also involves the same learning techniques and the same general experimental setting. The influence of the similarity dimensions such as we have previously discussed is kept to a minimum by making the tasks as dissimilar as possible while still being of the same class.

Briefly, let us consider quantitative indications of the extent of learning-how-to-learn transfer. Ward (399) required *Ss* to learn 16 successive lists of nonsense syllables, one list a day. Each list contained 12 syllables, and they were probably of roughly equal difficulty, although deviations in the practice curve may well be accounted for by some differences in difficulty. Values estimated from Ward's graph show that on the average, the first list required about 38 presentations to learn to a criterion of one perfect recitation; the fifth list in the series required 24 trials; the tenth list 18 trials, and the sixteenth list, 16 trials. Clearly, the gain is very large from the first list to the sixteenth, with the biggest improvement made early in the series, so that the general curve is much like the usual inverse negatively accelerated learning curve.

Schieber (333) has investigated the practice effects in learning successive lists of paired associates. Each list consisted of 10 pairs

of two-syllable adjectives. A different list was learned each day for four consecutive days, all *Ss* being naïve in paired-associate learning and in laboratory experience of any kind. The lists were not known to be of equal difficulty so that by a method of counterbalancing each list appeared an equal number of times on each of the four days. Schieber's results indicate that the practice effect in paired-associate learning of adjectives is not large. Her results for 24 *Ss* are shown in Fig. 51 in terms of mean trials to learn to a criterion of 8 out of 10 correct responses on each day. The only learning-how-to-learn transfer of any consequence occurs between the first and second days and the difference here is barely significant statistically. Explanation of this lack of learning-how-to-learn may lie in *Ss*' histories. It is possible that students, having to master foreign language vocabularies, having to learn definitions, and having to make symbolic translations of many kinds in their usual school work are actually quite well practiced on paired-associate learning (as compared with serial learning) before coming to the laboratory.

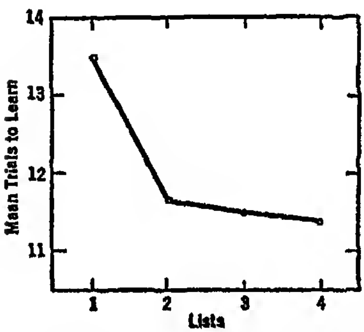


FIG. 51. Learning-how-to-learn paired-associate lists. Data from Schieber (333).

Maze learning shows a large practice effect. In an unpublished study by the writer 40 *Ss* learned two mazes in a single experimental session. Half the *Ss* learned Maze *A* first and the other half, Maze *B* first; hence, differences in speed of learning the second maze as compared with the first cannot be attributed to differences in the difficulty of the mazes. On the average, *Ss* took 20 trials to learn the first maze and only 11 to learn the second. There was no specific similarity relationship between the two mazes to account for the great improvement from first to second maze, so we must attribute the improvement to learning-how-to-learn transfer.

Many other data in the literature show that learning how to learn is one of the major sources of positive transfer. Just what

processes are involved in learning how to learn is unknown. Several factors are believed to be important: (1) *S* gets adjusted to the situation; this implies that he is less distracted by irrelevant stimuli; (2) unfruitful modes of attack on the problem may drop out when they prove unsuccessful; (3) correct associations in and of themselves may be formed more rapidly. These factors describe what is believed to take place, but of course they are not acceptable explanations as yet. The fact that the experimental situation remains constant from day to day, with only the specific material changing, suggests that the practice effects represent a gross differential reinforcement process wherein irrelevant responses to the situation are inhibited, and relevant responses brought to the fore.

Implications of learning-how-to-learn transfer. The importance of these practice effects is twofold: (1) They represent one of the major causes of positive transfer. This fact is reason enough for persistent caution against comparing amount of transfer in one situation with that in another when the practice effects for the two situations are not equal. Because of these practice effects we have indicated that Fig. 50 will probably not hold unless the *S*s are well-practiced. Practice effects will counteract conditions making for negative transfer and may well obscure any sign of it.

(2) The second important consideration regarding practice effects concerns experimental design. We have considered this problem only in regard to an individual's practice effect and have seen how such practice effect may be distributed equally over all conditions when the effect is not itself being studied. Of greater importance and of greater complexity are the problems involved when we counterbalance the practice effect for a group of *S*s taken as a whole. These will be taken up in detail in the following chapter.

Bi-Lateral Transfer

This form of transfer is usually positive. Typically, it has referred only to cases in which a motor skill developed through the use of one arm or leg is transferred to the arm or leg on the *opposite* side of the body. The transfer, however, need not be lateral; if a skill is developed by the left arm or hand, there will

also be positive transfer in acquiring the same skill with the left leg or foot.

A moment's reflection will demonstrate that bi-lateral transfer has been evident in your own experiences. The fact that you can write with your non-preferred hand, even though barely legibly, indicates that the long training with the other hand has had its effect.

One of the traditional laboratory tasks used in studying bi-lateral transfer is mirror star-tracing. A printed star is placed on a table before an upright mirror. A shield is attached to a rod several inches above the star in such a position that when *S* faces the mirror, he cannot see the star directly, but only in the mirror. *S*'s task is to trace around the outline of the star by directing his pencil movements from the mirror image of star, hand, and pencil. The shield is high enough to allow complete freedom of hand movement.

The outline of the star is customarily drawn with double lines, like a narrow pathway. In tracing the outline *S* must stay in the pathway. A complete circuit of the star constitutes one trial and the basic response measure is time taken to complete the circuit. Errors—number of times *S* leaves the pathway—may also be tabulated. On the first trial or two, old or normal movement habits produce considerable interference with movements required to follow the reversal of directions created by the mirror image. Bi-lateral transfer is demonstrated when it is shown that practice at mirror tracing with one hand transfers to performance with the other hand.

An experimental illustration. The results of an experiment performed in the writer's laboratory course may be used to illustrate the method and results of a simple bi-lateral transfer investigation. The students, after reading about bi-lateral transfer, were in considerable disagreement as to "what caused it." Some felt that it was due entirely to the learning of a general principle, the principle of mirror reversal, and that if one learned this, transfer would be complete from one hand to the other. Other students believed a certain amount of skill was required which would be specific to a given hand and that this would not transfer to the other. Accepting the fact that bi-lateral transfer occurs, no student suggested the hypothesis that acquisition

of a habit by one hand would not transfer at all to the other. However, this third possibility could be set up as an hypothesis along with the other two, so three hypotheses were stated:

1. In learning mirror star-tracing one learns only the general principle of mirror reversal. Hence, transfer from one hand to the other should be 100 per cent.

2. In learning mirror star-tracing one learns not only a general principle but also skills specific to the muscle groups of a single hand. Bi-lateral transfer will thus take place but it will not be 100 per cent.

3. In learning mirror star-tracing one learns a skill specific to the muscle groups of a given hand. No bi-lateral transfer will take place.

An experiment was designed to test these hypotheses. Three groups were required, one control and two experimental. All *S*s were given a preliminary practice trial, and the second trial was used as a source of data for matching three groups of 14 *S*s each. The matching was done on the basis of means and variability. One group was used for each of the following conditions:

	Trial 2 (Matching Trial)	*Trials 3 to 12*	*Trials 13 to 15*
Group I	Preferred Hand	No Practice	Preferred
Group II	Preferred Hand	Non-Preferred Hand	Preferred
Group III	Preferred Hand	Preferred Hand	Preferred

The critical point is the post test, trials 13 to 15. If hypothesis 1 is substantiated, the difference in performance on the post test should be no different for groups II and III and both would give higher scores than I. If hypothesis 2 is substantiated, performance of Group III should be better than that of Group II, and Group II better than Group I. If hypothesis 3 is substantiated, Group II should do no better than Group I on the post test trials and Group III should do much better than either I or II; hence, no bi-lateral transfer.

Figure 52 shows that hypothesis 2 is substantiated. As a result of trials 3 to 12 on the non-preferred hand, Group II learned something which aided performance on the test trials; otherwise, they would have done no better than Group I on test trials. This is evidence for bi-lateral transfer. Clearly, however, Group III

learned something which Group II did not, thus suggesting that certain skills specific to the hand are acquired in addition to the general principle of mirror reversal.

Extensive analyses of the mirror tracing situation (33, 65) have shown that transfer will take place from hands to feet, from feet to hands, on the opposite or on the same side of the body. Examina-

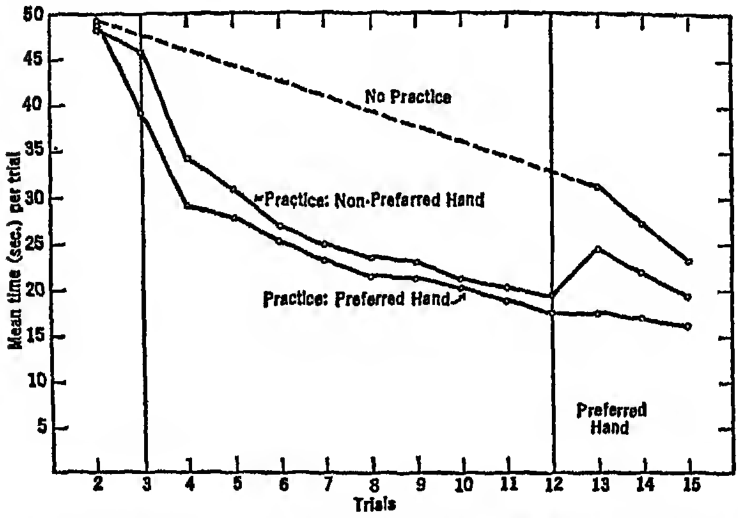


FIG. 52. Transfer in mirror star-tracing. Trial 1 was a practice trial: trial 2 was used as a source of data for matching the three groups of 14 *Ss* each. Tracing was with the preferred hand on trial 2.

tion of the situation in which such transfer is measured shows that both the stimulus situations and the responses required have high similarity so that positive transfer would be expected. Aside from such gross comparisons, however, the similarity relationships remain unanalyzed.

Related phenomena. A form of bi-lateral transfer appears to occur in certain conditioning situations. In one experiment (110) *S* was conditioned to withdraw his *right* hand at the sound of a buzzer. This was accomplished by giving several trials in which the buzzer was followed by shock to the hand. After the response was established, *E* placed *S*'s *left* hand on the electrodes and sounded the buzzer. On this transfer test 62 per cent of the *Ss*

withdrew their left hands although the left hand had never been shocked nor had it ever been placed in position on the electrodes before.

Another related form of transfer has been demonstrated by Wickens (407). *S*s were conditioned to withdraw a finger to the sound of a buzzer. During training the hand was strapped to a board, palm down, with only the finger free. The response was thus an *extensor response*. On test trials the hand was turned over so that the palm was up. The buzzer was presented on the test trials and transfer was 100 per cent; all *S*s responded with a *flexion response* of the finger, although again, this response had never been made during training.

In both of these instances, the fact that transfer takes place indicates that the response learned to the stimulus is more than an association between a buzzer and a discrete muscle group. What appears to be more nearly the case is that an association is formed between a stimulus and withdrawing, and such a response can be readily transferred. The likely mechanism involved is response generalization.

Negative Transfer in Animal Learning

The *A-B*, *A-C* paradigm in verbal learning usually results in negative transfer. Two illustrations of its use in animal learning will demonstrate that the negative effect is not as great as it is in human learning.

Foreleg flexion. An experiment by Kellogg and Walker (192) shows that when the *A-B*, *A-C* procedure is used with dogs the over-all transfer effects are positive, but the behavior of the animals also indicates the operation of negative processes.

Dogs were first trained to flex the right forepaw at the sound of a buzzer, thus avoiding a shock. After the response was well-established, the shock was transferred to the left forepaw, but the buzzer was retained as the stimulus. Then the dogs had to learn to lift the left paw to the sound of the buzzer. The criterion of learning was 19 avoidance responses out of 20 buzzer presentations. During learning of right foot flexion, the mean number of trials for 6 *S*s to reach the criterion was 154. When the shock was transferred to the left foot, an average of only 71 trials was required to reach the criterion. These figures indicate high posi-

tive transfer if we assume that, had the left foot been conditioned originally, about the same number of trials would have been taken to reach the criterion as were needed in conditioning right foot flexion.

From the description given of the animal's behavior there is considerable evidence of negative transfer during the early stages of acquiring left foot flexion. In many cases the dogs would lift first the right foot and then the left. Throughout the entire course of left foot-conditioning some of the animals lifted both legs simultaneously, a feat they were able to accomplish by supporting their body weight on a neck-stock. Thus, while the net effects of the training procedures indicate positive transfer, evidence of negative processes was apparent in the persisting right foot flexion.

T-maze learning. Several studies have shown that interference is present when rats are first trained to take one arm of the T then given training to break the habit and take the other arm For example, Bunch (48) used a water maze shaped like a T Rats will avoid water when possible, so the motivation was that of escaping water. By providing a platform at the end of the right arm onto which the rats could crawl, *E* initially taught the animals to make a right turn after swimming through the maze. The platform opened into a small box in which warm air was circulating. Such an escape was not possible at the end of the left arm of the maze. Successive trials were given until the rat had made 10 consecutive errorless "swims." In breaking the right turning habit and substituting the left turning habit, *E* moved the platform to the end of the left arm, thus making escape from the water impossible if the rat continued his right turning habit Successive trials were given until the criterion of 10 successive errorless trials was made. A control group learned only the left turning habit.

The transfer effects of learning the right turning habit were shown by comparing the performance of the experimental and control group in learning the left turning habit. The control group required a mean of 22.17 trials to learn the left turning habit; the experimental group required 24.80. In terms of errors (taking the "wrong" arm) the negative transfer effect is shown by the fact that the control group averaged 5.30 errors, whereas

the experimental groups averaged 7.87 errors. Negative transfer in learning the second habit as a consequence of prior learning of an antagonistic habit is clearly present.

Transfer in an Industrial Task

To allay any misconception that all transfer experiments have been done with nonsense syllables, buzzers, mirrors, and flexion responses, we should have one illustration of transfer in a somewhat more utilitarian setting.

Woodward (421) was able to find two industrial assembly tasks for which the movements were almost identical, although the materials being assembled were somewhat different. In terms of our analysis, the situation is comparable to an *A-B*, *A3-B* paradigm, in which the responses are almost identical and the stimulus situation only slightly similar.

Ss were girls, varying in age from 14 to 18 years, and all were attending vocational high school. A control group consisted of 48 *Ss*, an experimental group of 42. The design of the experiment is as follows:

	Pre-Test on Assembly 1	*Training on Assembly 2*	*Post Test on Assembly 1*
Control:	6 assemblies	————	12 assemblies
Experimental:	6 assemblies	57 assemblies	12 assemblies

The two groups differed only in that the experimental group was given 57 practice trials in constructing Assembly 2. The two groups were shown to be well matched on the pre-test trials so that differences which occurred on the post test were a function of the 57 training trials on Assembly 2. The transfer was positive. The two groups did not differ on the initial post-test trials, but by the end of the twelfth post-test trial the experimental group was performing significantly better than the control group.

Summary

In this section we have considered a group of transfer phenomena in which certain similarity relationships are present and are presumed to account for the obtained transfer. However, these studies have shown only that transfer occurs; they have not been concerned with variations in similarity dimensions.

SECONDARY VARIABLES INFLUENCING AMOUNT OF TRANSFER

We now turn to three variables which may cause the amount of transfer to vary. These are secondary in the sense that a certain degree of similarity must be present or these variables will have little influence. Given this unknown minimum amount of similarity, however, the influence of these secondary variables may be studied.

The three variables are: (1) degree of first task learning; (2) number of previous tasks; and (3) time between tasks.

Degree of First Task Learning

The findings concerning this variable are quite consistent: the greater the degree of first list learning the greater the probability of obtaining positive transfer and the less the probability of obtaining negative transfer. Such a statement will include the relationship for all ways of varying similarity which we have considered.

Experimental illustration. In one study, *S*s learned lists of paired associate two-syllable adjectives (394). Each list consisted of 10 pairs, and the similarity relationship between the two lists was *A-B, A-C.* Four conditions in the experiment were different only in terms of the degree of learning attained on the first task before going to the second. The degree of learning of the first list was controlled by setting different levels of mastery to be attained on each condition. The four conditions were as follows:

Condition	*Criterion of Learning: First List*	*Average Number of Trials Required*	*Criterion of Learning: Second List*
I (control)	———	———	6 out of 10 correct
II	3 out of 10 correct	4	6 out of 10 correct
III	8 out of 10 correct	11	6 out of 10 correct
IV	10 out of 10 correct plus 5 trials	20	6 out of 10 correct

The influence of these different degrees of learning was measured in terms of rate of learning the second list to a criterion of 6 out of 10 responses correct. Analyses of the results were made in terms

of correct responses on the initial learning trials of the second list, as well as in terms of number of trials required to reach the criterion. Such an analysis is necessary if differences in character of transfer as a function of stage of learning the second list are to be discovered.

The performance on the initial stages of learning the second list is shown in Fig. 53. The points are the mean correct responses

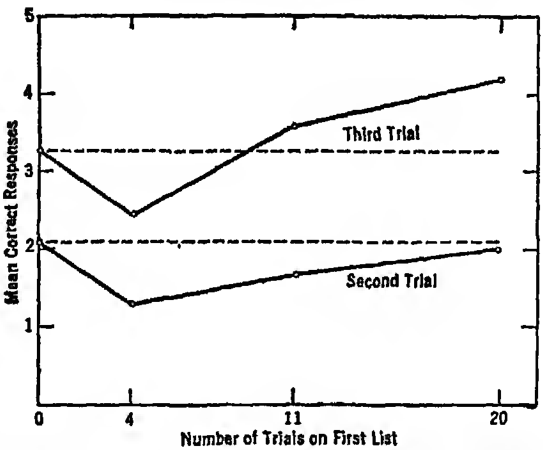


FIG. 53. Transfer as a function of degree of learning of first task. The curves show the mean correct anticipations on the first two anticipation trials (Trials 2 and 3) of the second list. Values below the dotted line for each trial represent negative transfer, those above positive transfer (394).

of 48 *Ss* for each degree of first list learning on the first two anticipation trials. The first two anticipation trials are actually trials 2 and 3, since the first trial is always a study trial, in which *S* observes the list and tries to form associations. The first attempted anticipations of the correct responses take place on the second trial.

Figure 53 shows clearly that with a low degree of first list learning, negative transfer is produced on the first two trials. The *Ss* in this experiment were well-practiced in learning paired associates so that very little learning-how-to-learn transfer is present to obscure negative effects in learning *A-C* following learning of *A-B*. With high degrees of learning of the first list no negative transfer is evident on the second trial, and by the third trial

positive transfer has occurred—a change in one trial from zero transfer to positive transfer.

The *relative* relationships of the transfer conditions hold fairly well in terms of trials to reach the criterion on the second list. In terms of mean trials the values are 6.71, 6.91, 5.57, and 4.98 for Conditions I through IV respectively. Note, however, that from these means we have no evidence of negative transfer. The difference between Conditions I and II suggests the operation of negative transfer but it is far from statistically significant. Had the learning of the second list been to a criterion of one perfect trial, all experimental conditions probably would have shown positive transfer. These data emphasize that in studying transfer, all stages of second list learning must be analyzed if the true picture of transfer effects is to be obtained.

The relationship analyzed. The general principle we have stated is that the higher the degree of first task learning the greater is the probability of obtaining positive transfer and the less is the probability of obtaining negative transfer. Such a relationship is difficult for some students to understand. They believe that under high degrees of first task learning the response tendencies become so strong that they will persistently interfere with the learning of the second task. Yet, if we recall our study of conflict we will remember that greatest conflict occurs when response tendencies are of about equal strength. With low degrees of first task learning we would thus expect the greatest interference to occur early in the learning of the second task when the response tendencies of this list also are weak. This is exactly what the data show.

On the other hand, with high degrees of learning of the first task we would predict the interference to be maximal fairly late in the learning of this second task, i.e., when those response tendencies become nearly equal in strength to those of the first list. Fragmentary evidence tends to support this prediction, although it is very difficult to focalize the interference at one specific point late in second list learning.

In addition to factors making for specific interference between items, we have three factors which seem to be important in determining transfer as a function of degree of first task learning.

1. *Learning-how-to-learn transfer.* We know that the greater the number of trials on the first task, the greater is the amount

of learning-how-to-learn transfer. This process was probably not important in the study reported above since *S*s were already well practiced.

2. *Transfer of differentiation.* From Gibson's (106) careful analysis there is reason to believe that a transfer of differentiation takes place in certain situations. Let us explain what we mean by differentiation and its transfer. In learning the first list, *A-B,* the stimulus items may have some similarity among themselves. This would tend to produce generalization which would inhibit learning. By differential reinforcement the generalizing tendencies are reduced among these stimuli as learning progresses. Now, if the second list is *A-C,* the learning process will not require this differentiation process since it has already taken place among the same stimuli during learning of the first list. Furthermore, we would expect that the greater the number of first list trials the better would be the differentiation; hence, the greater would be the positive transfer in learning the second list.

There is some independent evidence that such positive transfer is plausible (117, 390). This transfer could be expected only if the stimuli were identical or highly similar. It is possible that a similar transfer may take place as a result of response differentiation. If so, it would tend to produce positive transfer in learning *A-B, C-B.* The evidence for transfer of differentiation is such that we should accept it only tentatively as a cause of positive transfer.

3. *Gross differentiation.* Finally, there is a possibility that differentiation may take place at a gross level. *S,* after having learned the first list to a high degree, may be able to discriminate clearly that these responses are "wrong" when he starts to learn the second list. Descriptively, it is as if *S* is able to dissociate these first list responses more rapidly and go about learning the correct responses—the second list responses. Such a process would not make for positive transfer; it would tend only to reduce the amount of negative transfer and would be directly related to the degree of first list learning.

Number of Previous Tasks

In all our illustrations thus far we have considered the influence of *one* task only on the learning of a second task. Now we inquire

into the relationships of direction and amount of transfer as a function of two or more tasks.

Experimental illustration. In one experiment (390) *S*s were presented 0, 2, 4, or 6 *lists* followed by another list which was presented until *S*s correctly anticipated 6 or more responses out of the possible 10. Each variation in the number of lists called for a separate experimental condition. All were paired-associate lists of adjectives, and for a given condition, all lists had the same stimuli but different responses. Thus, in the condition in which 6 lists were learned before the final common list, *S* might have actually attached seven responses to the same stimulus.

The lists were presented for four trials each, with *S* instructed to get as many correct responses as possible on each trial. An interval of 45 seconds intervened between each list. Two sets of data may be obtained from this experiment: (1) the transfer effects in going from the first through the sixth list (these lists appeared equally often at each position, so differences in the difficulty of the lists would be equalized for each stage); (2) the effect of learning 0, 2, 4, and 6 lists, for 4 trials each, on the learning of the final common list to a criterion of 6 out of 10 correct responses. We will consider these results in order.

In calculating the results of the performance on the six lists, we may actually consider the influence of each list on the successive lists by presenting them (in order) along the base line of a graph and by placing the mean correct responses obtained on each of the three anticipation trials along the ordinate. These results are shown in Fig. 54.

Figure 54 shows that negative transfer was produced in the learning of all lists except, of course, the first. It is further evident that as learning proceeds from trial two through trial four, the greater the number of previous lists the more rapid is the dissipation of the negative transfer. That is, by the fourth trial the performance on lists 4, 5, and 6 indicate positive transfer whereas this is not true with lists 2 and 3.

Figure 54 indicates that by the fourth trial positive transfer effects are evident if three or more lists have been previously learned. This tendency is very clear-cut when we consider the mean number of trials to learn the seventh list to a criterion of 6 out of 10 correct responses. With no previous lists (control), the

common list was learned to the criterion in an average of 6.49 trials; with 2 previous lists, 6.08 trials were required; with 4 previous lists, 5.67 trials, and with 6 previous lists, only 5.38. Had the learning of the common list been carried to a higher level of

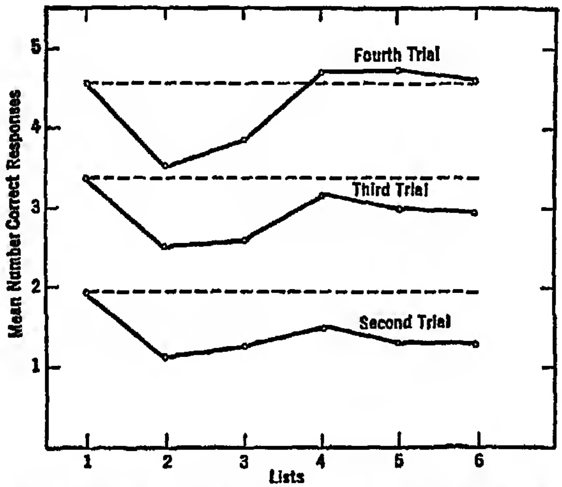


FIG. 54. Transfer in learning six successive lists of paired-associates, where each list had the same stimuli but different responses. The dotted lines indicate zero transfer effects and all points below these lines are indicative of negative transfer. Each list was presented for four trials. N is 72 for the first two lists; 48 for lists 3 and 4, and 24 for lists 5 and 6 (390).

mastery it is very likely that differences among the conditions would have been still greater. On the basis of the data available, we are able to state the general principle that as the number of prior or previous lists increases, the greater is the amount of positive transfer which eventually occurs.

Time Between Tasks

This variable concerns the relationship between amount and character of transfer and the time intervening between the two tasks. In considering this variable we must evaluate two separate aspects of transfer. First, we have transfer of the learning-how-to-learn type. This phenomenon is largely dependent only on the learning of successive samples of the same class of materials.

Secondly, we must evaluate the effects of the time interval between first and second tasks in situations where the direction and extent of transfer is due to specific similarity dimensions.

Time and learning-how-to-learn. Facts concerning learning-how-to-learn transfer and the interval between tasks are quite consistent for human *S*s. Several experiments by Bunch and his students have shown that practice effects are independent of the time between tasks, i.e., the practice effects remain constant irrespective of interval length. Briefly, let us illustrate this conclusion.

In two separate experiments (50) different groups of *S*s learned paired nonsense syllables and then learned a second list of unrelated syllables after intervals up to 28 days. A control group learned only the second list. Compared with the control group, the experimental groups learned the second list in about 40 per cent fewer trials. This was true of the group that learned the second list immediately after the first and of the group that learned the second list 28 days after the first.

Bunch's work with rats (e.g., 49), has suggested that as the time interval between mastery of the two mazes increases, the amount of positive transfer first increases and then slowly decreases, but it is still as great after 120 days as after 1 day.

How do we account for the fact that there is no loss in learning-how-to-learn transfer with the passage of time? This is a difficult question. We have already indicated that the processes which produce practice effects are obscure; we describe it simply by saying *S* learns how to learn. Bunch's results show that unlike the content which *S* learns, the techniques by which he learns the content are not forgotten. We believe that most forgetting is caused by interference from other activities. If we assume that learning-how-to-learn represents a normal learning process, we must conclude tentatively that techniques of learning a given task are not susceptible to associative interference from other techniques or activities.

Similarity and the time interval. There are no data from adequately designed experiments which give us information on this relationship. That is, given a situation in human learning which produces negative transfer when the second task comes immediately after the first, we do not know how the transfer changes

as a function of the time interval between the two tasks. Furthermore, the situation is too complex to attempt a prediction.

SUMMARY

1. Transfer of training defines the area of study in which *E* determines the influence of one activity on subsequent performance of another task.

a. If the influence of one activity is that of facilitating performance on another, the transfer effect is called positive.
b. If the influence of one activity inhibits or retards performance on another, the effect is called negative.

2. We have continually emphasized that processes producing positive transfer and those producing negative transfer interact and that our measurement at a single point in performance on the second task represents the net effect of the two processes. Analytical experiments will examine transfer effects at various stages of learning the second task.

3. Similarity is the one major variable determining transfer effects. We have used the paired-associate learning technique as an analytical tool for predicting the influence of different similarity relationships between the two tasks. It was our hypothesis that similarity determines transfer because of its effect on generalization. It is through the study of similarity that we are able to discern why an organism makes the responses it does when placed in a problem situation.

4. Dimensions of similarity which have been shown to influence transfer were illustrated. These dimensions are:

a. Formal identity
b. Formal similarity of figural designs
c. Meaningful similarity

A graph for predicting transfer effects in paired-associate type learning summarizes this section.

5. Transfer effects based on similarity which has not been dimensionalized were illustrated. These include:

a. Learning-how-to-learn transfer
b. Bilateral transfer and related phenomena
c. Negative transfer in animal learning
d. Transfer in an industrial situation

6. Secondary variables influencing amount of transfer are:

   a. Degree of first task learning
   b. Number of previous tasks
   c. Time between tasks

7. Methodological implications of transfer of training will be included in the next chapter on experimental design.

CHAPTER X

Design Methods III and IV

INTRODUCTION

In the previous chapter we did not discuss the problems of experimental design associated with the study of transfer of training. Because of the special importance of these design problems it was felt that a single chapter should be devoted to them. As compared with Design Methods I and II, Methods III and IV are considerably more intricate. Much of the experimentation to be reported in later chapters has been done by Design Methods III and IV so it will be well to get a good understanding of them at this point in order that the subject matter sequence will not have to be disrupted too greatly.

In discussing Design Methods I and II in Chapter V, we indicated that both methods involve the use of a different group of *S*s for each condition of the experiment. *Design Methods III and IV use a single group of Ss for all conditions; each S in the group serves in all conditions.*

DESIGN METHOD III. COUNTERBALANCING OF CONDITIONS

In the previous chapter we have seen that when *S* learns successive samples of the same class of material, there is a practice effect—*S* learns how to learn. In Design Method III, *S* serves in all conditions. This means that if we have a two-condition experiment, the second condition will be influenced by a practice effect, whereas the first will not. Since a good estimate of the relative influence of two stimulus conditions cannot be obtained when one is influenced by practice effect and the other is not,

we need some way to equalize the practice effects for both conditions. This may be handled by the counterbalancing of conditions.

Equalizing of practice effects by counterbalancing. In Chapter II we pointed out how practice effects could be equalized for a given *S* in all conditions by the use of the *abba* order of conditions. Now suppose we are doing an experiment in which *S* learns a list of words on each of two conditions, *A* and *B*. We might suggest that to balance practice effects we could present the conditions in the *ABBA* order; that is, we could use four lists and require each *S* to serve in each condition twice.

If the curve of learning-how-to-learn were linear, i.e., if each successive list resulted in an equal practice effect, *ABBA* would be satisfactory. However, the curve is not linear. When we use the *abba* order in psychophysical experiments, the curve need not be linear for we can break the blocks of trials into many *abba* orders so that non-linearity of the curve will not appreciably influence one condition more than the other. In the case of lists of words, however, we could not shift from one list to the other after each trial; instead we would have to finish each condition by learning the entire list. If the practice curve is like Fig. 51, Chapter IX, and we applied *ABBA*, the *A*s and *B*s would certainly not be equal with respect to practice effects; *A* would be influenced more by practice than would *B*.

One way of handling this problem is to put half the *S*s through the *ABBA* order and the other half through the *BAAB* order. If the two groups produced about the same practice curve, the *A*s and *B*s would be equally influenced by practice effects. It is possible, of course, that the two sub-groups might not show the same practice curve, though such a happening is not too likely, and the error resulting from it would not be great.

Another method of handling the problem is to practice the *S*s well before starting the experimental conditions. Once the practice curve becomes nearly horizontal, it is almost linear and we can safely use the *ABBA* order.

In experimental work neither of the above methods has been used frequently. Instead, a method which balances practice effects among *S*s rather than for each *S* has been employed. It is this method which we are calling *Design Method III.*

With the use of Design Method III, and two conditions, half the *S*s take condition *A* followed by *B*, and half take *B* followed by *A*. In such a procedure each condition occurs equally often at each stage of practice; hence the influence of the practice effects will be the same for both conditions.

In counterbalancing among *S*s, each condition not only occurs equally often at each stage of practice, but also, precedes and follows all other conditions an equal number of times. If we have two conditions, *A* and *B*, there are only two possible ways in which the two can be arranged in order: *A* to *B*, and *B* to *A*. Thus, in counterbalancing an experiment with these two conditions our plan would appear as follows:

	1st Session	*2nd Session*
Subject No. 1	A	B
Subject No. 2	B	A

Since we would not normally use just 2 *S*s in an experiment, we would continue this plan for successive *S*s, so that one half of our group would take the *A* to *B* order, and the other half the *B* to *A* order. We would assume that the responses at the second session would be more greatly influenced by practice effect than would the responses at the first session. However, since both conditions would have occurred equally often on the second session, no differential bias in the responses for the two conditions would be present. Note that counterbalancing does not eliminate practice effects; counterbalancing only distributes these practice effects equally over all conditions when the effects are considered for all *S*s combined.

The above procedure rests on one assumption; namely, that the practice effects for both groups are roughly the same. Probably the safest procedure is to give at least two practice days to check on the differences in magnitude of the practice effects between the two groups before giving the experimental conditions. If we give several days' practice before the experimental conditions so that the practice curve for both groups is fairly level, the two groups might be quite different in ability, and still the procedure would be sound. The reason for this will appear shortly. Actually, when we use several conditions, so many different orders are used that it would be relatively meaningless to speak of different practice

curves; hence, a bias as a result of different curves is extremely unlikely. Since practice days are usually given before the experimental sessions, we shall probably produce an experiment in which the practice effects which are present are equally distributed over all conditions.

In the statistical handling of the data, the performance of a given *S* on one condition is compared directly with his performance on the other. The precision of the method is gained from the fairly sound assumption that the best match for a given *S* is that *S* himself. We shall point out later, however, that there are certain situations in which it is not advisable to use the method. One of these must be mentioned at this point. If *E* has reason to believe that the effect of going from *A* to *B* is quite different from the effect of going from *B* to *A*, the method should not be used since it would give a distorted picture of the influence of the experimental conditions as such.

A three-condition experiment. Now, let us assume we are going to run an experiment in which there are three experimental conditions, *A*, *B*, and *C*. To follow the rules of counterbalancing, we must see that each condition occurs equally often at each stage of practice, and that it precedes and follows all other conditions an equal number of times. For three conditions, the number of possible orders of conditions is six:

	Session 1	*Session 2*	*Session 3*
S No. 1	*A*	*B*	*C*
S No. 2	*A*	*C*	*B*
S No. 3	*B*	*A*	*C*
S No. 4	*B*	*C*	*A*
S No. 5	*C*	*A*	*B*
S No. 6	*C*	*B*	*A*

There is no other combination of orders in which the three conditions may appear. We note that *A* occurs twice at each session, as do *B* and *C*. Again, we probably would not use only 6 *S*s; rather, we would use, let us say, 30, thus repeating the above block five times. As long as number of *S*s used is a multiple of six we have perfect counterbalancing of three conditions.

Number of conditions and number of *S*s. With two conditions we have seen that 2 *S*s are required to effect complete counterbalancing. With three conditions, 6 *S*s are required. How many

Ss are required to have complete counterbalancing if we have four conditions? If you write all the possible combinations, you will find that 24 *Ss* are required for a four-condition experiment. With five conditions 120 *Ss* would be required—120 *Ss* with five conditions each. Seldom does *E* run such an extensive number of *Ss* in so many conditions. Consequently, no more than four conditions are customarily used in counterbalanced designs. With five or more conditions, Design Method IV is used. This method will be explained shortly.

Three Variations of Counterbalanced Design

In practice there are three general kinds of experiment which may employ counterbalancing: (1) experiments in which *E* varies the nature of the material; (2) experiments in which *E* varies conditions under which material is presented, and (3) those in which *E* studies practice effects. Each of the three kinds of experiments involves somewhat different procedures in setting up the design and in collating the data.

Variation in material. In these experiments *E* varies the nature of the material to be learned. This might mean variation in similarity, in meaningfulness, in difficulty, and so forth. Let us say we are going to perform an experiment on transfer of training as a function of similarity between two lists of nonsense syllables. We know the lists we are using vary in only one aspect—degree of similarity. Since each *S* is going to serve in all three conditions, we need three sets of two lists each. Set *A* has high similarity between the two lists; Set *B*, medium similarity; Set *C*, low similarity. Our design would be exactly the same as the three-condition experiment outlined above.

For simplicity, let us say we use only 6 *Ss*, and a single measure of transfer; namely, trials to learn the second list. For each *S* for each condition we obtain the following raw scores:

	Session 1 Cond. Score	*Session 2* Cond. Score	*Session 3* Cond. Score
No. 1	*A* — 10	*B* — 11	*C* — 8
No. 2	*A* — 14	*C* — 12	*B* — 13
No. 3	*B* — 16	*A* — 10	*C* — 7
No. 4	*B* — 15	*C* — 11	*A* — 8
No. 5	*C* — 11	*A* — 10	*B* — 12
No. 6	*C* — 7	*B* — 9	*A* — 4

Now let us collect the scores for the three conditions:

	A	*B*	*C*
No. 1	10	11	8
No. 2	14	13	12
No. 3	10	16	7
No. 4	8	15	11
No. 5	10	12	11
No. 6	4	9	7

From these distributions of raw scores we can determine directly whether or not the conditions produced significant mean differences in speed of learning. *A* can be compared with *B* and *C*, and *B* can be compared with *C*.

Variation in conditions. By this we mean that *E* varies certain environmental conditions, such as time factors, distribution of practice, amount of a given drug, and so forth. For our purposes, let us again consider a transfer experiment in which we vary the time between the presentations of the two lists. We use as our three intervals, 0, 30, and 60 minutes. We make up three sets of lists, each set containing two lists. They are so constructed that we believe they are roughly equivalent in difficulty. We arbitrarily assign two lists to be used as Set I on the first session—another two for Set II on the second session, and the final two to be used as Set III. Our Design for the three conditions now appear as follows:

	Session 1 *Set I*	*Session 2* *Set II*	*Session 3* *Set III*
S No. 1	0 min.	30 min.	60 min.
S No. 2	0 min.	60 min.	30 min.
S No. 3	30 min.	0 min.	60 min.
S No. 4	30 min.	60 min.	0 min.
S No. 5	60 min.	0 min.	30 min.
S No. 6	60 min.	30 min.	0 min.

Again we have each condition occurring equally often at each stage of practice so there should be no bias as a consequence of differential practice effects. We would bring our data together for the 0-minute condition, for the 30-minute condition, and for the 60-minute condition in a manner comparable to that used when varying similarity.

We have used Set-I lists on the first day for all *Ss*, Set II on the second day, and Set III on the third. These lists which we

constructed were only roughly equivalent—we cannot overlook the possibility that they may actually be significantly different in difficulty. At this point many students will object; they will insist that the lists have to be equal in difficulty if the same set is always being used on the same day. But, note, each condition occurs equally often with each set of words (or vice-versa) so that if the lists *do* differ in difficulty the effects of this variation will be distributed equally over all conditions; the difficulty variation will not, therefore, bias any given condition. In short, equalizing the effects of differences in difficulty by counterbalancing is precisely the same as equalizing practice effects. We shall point out later certain desirable features of fairly equal lists, but from the standpoint of sheer design the lists do not have to be known to be of equal difficulty.

Studying practice effects. Occasionally, there is an experiment in which *E* desires to study practice effects or, learning-how-to-learn transfer. If this is to be studied adequately, our design will be somewhat different from either of the above two. Again, we will consider three conditions, but this time our three conditions are *stages of practice*, Stage 1, Stage 2, and Stage 3. Arbitrarily we choose three lists of words, which we believe to be fairly equal in difficulty. Again, the lists do not have to be of equal difficulty. Now, what are we going to counterbalance? Our experimental variable is the stage of practice; obviously we can't counterbalance it; we can't give *S* the third stage of practice before he has had the first two stages. There is only one thing to counterbalance—the material, which, in this case, is the lists. Our design would appear as follows:

	Session 1 *Stage I*	*Session 2* *Stage II*	*Session 3* *Stage III*
S No. 1	List 1	List 2	List 3
S No. 2	List 1	List 3	List 2
S No. 3	List 2	List 1	List 3
S No. 4	List 2	List 3	List 1
S No. 5	List 3	List 1	List 2
S No. 6	List 3	List 2	List 1

Since we do not know that our lists are of equal difficulty, the counterbalancing results in each list occurring equally often at each stage of practice. Consequently, differences in rate of

learning from Stage 1 to Stage 2 to Stage 3 must be a function of learning how to learn, not a function of the lists. If we had three lists known to be of equal difficulty, no counterbalancing of any kind would be necessary.

The collection of data in this experiment would be different from that in either of the two preceding types. Here we are studying rate of learning at a given stage of practice; hence, we would compare the scores obtained on Session 1 with those on Session 2, and those on Session 2 with those on Session 3.

In summary of these three types of experiments using counterbalancing, we may say that if we are manipulating type of material, the material is counterbalanced; if we are varying an environmental condition, the material is held constant from day to day and our counterbalancing pertains to the environmental condition. Finally, if we are studying practice effects, we counterbalance the material. Our illustrations have used three conditions, but all considerations apply in exactly the same manner for two or four conditions.

DESIGN METHOD IV: SYSTEMATIC RANDOMIZATION OF CONDITIONS

We have said that if we have five or more conditions, complete counterbalancing requires more *Ss* than *E* may be able to run in a single experiment. *Systematic randomization*, so-called, has been devised for experiments in which five or more conditions are desired, and in which all *Ss* are to serve in all conditions.

The method. In point of fact, the label, systematic randomization, is something of a contradiction in terms because we cannot have randomization if there is a system. The method has three characteristics:

1. Each condition occurs equally often at each stage of practice.
2. Each condition precedes and follows all other conditions an equal (or approximately equal) number of times.
3. All possible orders of conditions are *not* used.

System is involved in that we make sure each condition occurs equally often at each stage of practice, and that each condition precedes and follows each other about an equal number of times. Other than meeting these requirements, the order of conditions

is random, although in actual practice there is less random order involved than might appear at first.

The method does not necessarily suffer because all possible orders of conditions are not used, nor because each condition does not precede and follow each other an equal number of times. Like counterbalancing, systematic randomization should not be used if there is reason to believe that one condition transfers to a second condition more than the second does to the first. Differential transfer effects would result in biased estimates of the influence of conditions; since we do not use systematic randomization if there is differential transfer, we have no cause for concern over the fact that all orders of conditions are not used. We may, of course, analyze the results of an experiment to determine if there have been differential transfer effects.

The number of *S*s needed to fill out the plot of systematically randomized conditions is a multiple of the number of conditions. If we have five conditions, we use 5, 10, 15, and so forth, *S*s and with six conditions, 6, 12, 18, 24, and so forth, *S*s. As an illustration, let us show how systematic randomization would be set up for six conditions (*A*, *B*, *C*, *D*, *E*, *F*) and six *S*s, again recognizing that we would seldom use only one *S* for each condition sequence.

	Session 1	*Session 2*	*Session 3*	*Session 4*	*Session 5*	*Session 6*
S No. 1	*A*	*B*	*F*	*C*	*E*	*D*
S No. 2	*D*	*E*	*C*	*F*	*B*	*A*
S No. 3	*C*	*A*	*E*	*B*	*D*	*F*
S No. 4	*F*	*D*	*B*	*E*	*A*	*C*
S No. 5	*B*	*C*	*D*	*A*	*F*	*E*
S No. 6	*E*	*F*	*A*	*D*	*C*	*B*

In this plot, each condition occurs once on each day, and each condition follows each other condition an equal number of times.

All considerations which we have given to counterbalancing designs with respect to types of experiments, collecting scores, and so forth, apply to systematic randomization in the same fashion as they do to counterbalancing.

Why wouldn't we use complete randomization of conditions? In a six-condition experiment it would seem, for example, that we might put six numbers in a hat and let the order in which these numbers are drawn indicate the order for one *S*. A second drawing would indicate the order for a second *S*, and so on. Like

Design Method I, no objection can be sustained against this method if a large number of *S*s is used, for in the long run we would expect each condition to occur about equally often at each stage of practice, so that each condition would precede and follow each other almost an equal number of times. But when we use small numbers of *S*s, let us say 20 to 50, as we almost inevitably will if we have five or more conditions for each *S*, we need not run the risk that chance will not equalize for practice effects. Therefore, we make certain that each condition occurs equally often at each stage of practice.

The Use of Practice Days

In many experiments it is customary to give *S* from one to several practice days before starting him on a sequence of experimental conditions. The idea behind the use of practice days is to get *S* adjusted to the experimental situation and to eliminate the sharpest portion of the practice curve—the curve of learning-how-to-learn. If the experimental conditions are somewhat complicated, it often appears desirable to introduce *S* to the various changes in the experiment on practice days so that he is better prepared for the experimental conditions when they occur. Such preparation is likely to reduce the variability of his performance, and reduction of variability results in a more reliable estimate of the influence of experimental conditions. We might expect a reduction in variability of performance only as long as the number of sessions involved is not great enough to set up a counteracting factor, such as boredom, which might increase the variability.

There is certainly no specific number of practice days which one can say should be given. When practice effects are great, more practice days are indicated than when practice effects are slight. We would recommend more practice days with serial learning than with paired-associate learning, because the learning-how-to-learn effects of serial learning are greater (see previous chapter). If the experiment is complicated, so that actually carrying out instructions might be difficult, practice days should be used to make sure that *S* understands what is expected of him.

The statistical value of practice days. As an exercise in experimental design, let us consider what statistical influence

practice days have on the scores for the experimental days. For an illustration, we will take an hypothetical two-condition counterbalanced experiment. The nature of the conditions is unimportant for our purpose. Let us assume that for the material being used

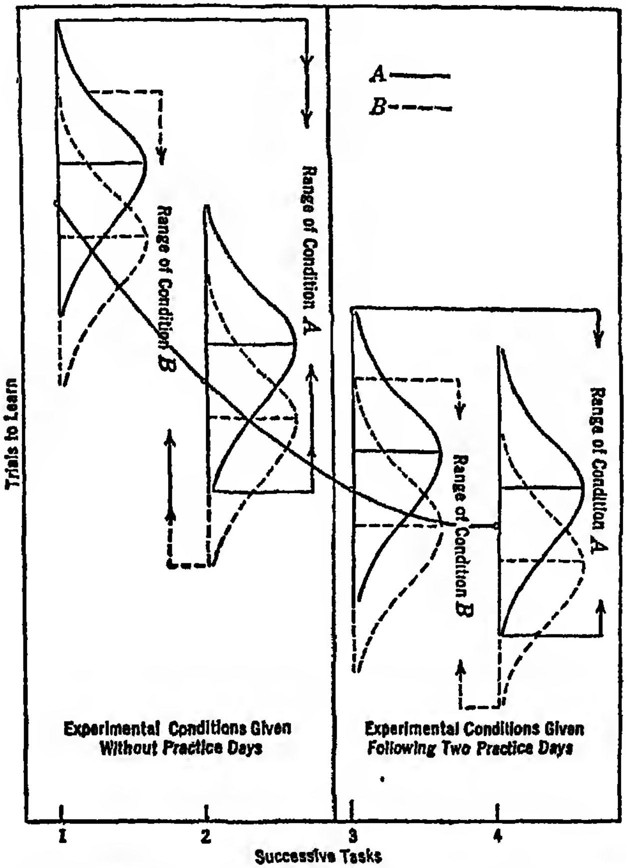


FIG. 55. The statistical influence of practice days. See text for explanation.

the learning-how-to-learn transfer is largely complete by the time *S* has learned two lists. This means that the sharp drop in the practice curve has taken place by the end of learning two lists and that for lists three and four the curve is nearly horizontal. This practice curve is shown in Fig. 55.

First, let us evaluate the situation in which we do not use practice days in our two-condition experiment. Our two conditions, *A* and *B*, are counterbalanced. This means that half the *Ss* get Condition *A* with List 1 and Condition *B* with List 2, and the other half get Condition *B* with List 1 and Condition *A* with List 2. The hypothetical distributions of scores for each condition on each list is shown in Fig. 55. The distribution of scores for Condition *A* is represented by the solid line and for Condition *B* by the dotted line. The distributions are so drawn that a difference in performance for the two conditions is apparent. Two assumptions are made in drawing these distributions: (1) the variable being investigated is uninfluenced by the stage of practice, i.e., the mean difference between Conditions *A* and *B* is the same for all stages of practice; (2) the range of scores remains constant at all stages of practice. (It is probable that the range actually decreases as a function of the practice stage. If this is true the validity of the following argument is enhanced.)

Now, let us see what happens when we list our scores for Conditions *A* and *B* when no practice days have been given. All the scores on Condition *A* we list together; we do the same with Condition *B*. This means that for each condition we are putting scores together from two rather different stages of practice. The resulting distribution for each condition would be that distance from the poorest score on List 1 to the best score on List 2. These score distances are depicted for you in Fig. 55.

As a second step we do exactly the same for the right half of the graph, in which the scores on the experimental conditions are depicted after having two practice lists. Again, we determine the ranges for Conditions *A* and *B*.

We may next note that the difference in the ranges of the conditions are quite different depending upon whether or not we have given practice days. The range of scores for a given condition is much greater when practice days *are not* given than it is for the same condition when practice days *are* given. What is the significance of this? We know that the standard error of a difference is directly proportional to the range; the greater the range the larger is the standard error, the smaller the range the smaller is the standard error, providing, of course, *N* is constant. With a constant mean difference between two conditions, the

critical ratio or t value is proportional to the size of the standard error of that difference, with the larger the standard error the smaller the critical ratio. We thus see that, given a constant mean difference, we increase the reliability of this difference if we have introduced practice days.

We have mentioned previously that in manipulating conditions under which a given material is learned, the materials used from day to day do not have to be of equal difficulty. But, we have also said it is desirable that such material be of near-equal difficulty. The reason for this, it should now become clear, is exactly the same as the reason for having practice days. If the material to be learned is unequal in difficulty, the range of scores for a given condition is increased. As an exercise, you should demonstrate this to your own satisfaction in a manner similar to that used in demonstrating the influence of practice days.

The Design Method to Use in Practice

There are no hard and fast criteria by which to choose a method for any particular kind of experiment. In many instances all methods would be equally satisfactory. The choice of the method sometimes hinges on the number of *S*s available, the amount of time to be spent per *S*, the number of conditions, and so on.

Because of the precision involved, Design Methods III and IV are preferred for a great many experiments. There are, however, certain instances in which these methods are not suitable because their use is likely to obscure the true influence of a given group of conditions. Fortunately, we can be fairly definite on our identification of these situations where it does not seem advisable to use Methods III and IV, but rather, Design Method II, Matched Groups. These situations fall into two classes:

1. Experiments in which there may be a considerable difference in transfer effect from one condition to another. Thus Condition *A* might strongly influence *B*, but *B* would have little effect on *A*. If this were the case, we would be getting a strongly distorted account of the influence of the condition being manipulated. Two general classes of experiments fall into this heading:

a. Experiments in which one or both conditions involve unusual treatment. Illustrations would be experiments in which electric shock was used or in which *S* was severely frustrated. It is very likely that the set or attitude

of *S* in going from a non-shock to a shock condition would be quite different from the set or attitude of an *S* going from a shock to a non-shock condition. Unless we are specifically studying the effects of differential sets, we do not want it to enter the experimental picture. Most of the experiments reported in the chapters on Motivation, Conflict, and Frustration were done by Design Method II, Matched Groups, rather than either Methods III or IV. *Es* have rightly believed that counterbalancing or systematic randomization would not give a true picture of the influence of the unusual treatments applied to their *Ss*.

b. Experiments in which the effect of varying instructions is to be determined. For example, we might want to know the influence on learning of rehearsal during a rest period. If in one condition we instructed *Ss* to rehearse, and in another condition did not instruct them at all concerning what to do during the rest interval, it is probable that counterbalancing would reduce the measurable influence of the conditions. *Ss* whose rehearsal instruction came on the first day might also rehearse during the second condition. If any experiment carries the likelihood that instructions intended for only one condition may carry over to another, Design Methods III and IV should not be used.

2. Experiments in which the task involved is such that one cannot obtain successive samples of that task. An illustration of this would be in the mirror star-tracing experiment. From the data presented it can be seen that learning is quite precipitous, and that *S* reaches a high level of performance after relatively few trials. In verbal learning, we could give *S* a completely different list on the next day; in mirror star-tracing, however, we cannot give him a star-tracer apparatus and expect his performance to start at a low level again. If we attempted to counterbalance conditions in mirror star-tracing, where each condition involved several trials, we would greatly diminish the influence of our conditions since *S* readily gets to a point where his performance will improve but little with practice. In short, when we have a task which is the only one of its class, or when the positive transfer to other members of its class is very considerable, counterbalancing or systematic randomization should not be used. Different groups of *Ss* have to be used for each condition.

These criteria are the only clear-cut ones which can be laid down in choice of method. One guide to selection is often available; if an experiment has been performed using a given technique, and if that technique seems sound, adapt it for your experiment.

Not only does this increase the probabilities of having a good experiment methodologically, but it also tends to integrate the procedures and results among different laboratories.

THE PROBLEM OF SCIENTIFIC GENERALIZATION

By now we are quite familiar with methods for designing sound experiments. We shall, of course, need additional practice in mastering the details of these methods as they apply to a given subject matter area, but we have the necessary tools to design and execute simple experiments and know that our data are sound. An experiment allows us to discover a factual relationship between stimulus and response variables. Now let us discuss briefly what we might say about the broader scope of this relationship, i.e., what we can say about this relationship in other situations with different *S*s.

Whenever *E* performs an experiment he usually does so with the idea that his findings will not be true *only* for his conditions and *only* for his sample of *S*s. The problem of scientific generalization is: how widely applicable are the results of the usual experiment? Many *E*s, fully cognizant of the dangers of unrestricted generalization, may cautiously state: "Under the conditions of the present experiment" such and such was found, and no direct attempt is made to assess the generality of the findings. However, this dodges a clear responsibility. One cannot expect to set up a specific experiment in order to derive facts for every aspect of all possible conditions for all people for every phenomenon. Sooner or later some generalization must be attempted. The human mind should never rest content with discovery of the purely specific.

There are four hierarchical factors which must be evaluated in considering the generality of an experimental finding: (1) the sample, and its representativeness of a given population; (2) the task or material used in the experiment; (3) specific conditions from which the facts were derived; (4) dimensional extremities.

In evaluating each of these factors, we must keep in mind at all times that it is assumed that *E* is also conversant with the pertinent findings of other *E*s and that these findings are considered in relation to his own.

The sample. Strictly speaking, if E has a random sample of a defined population he is justified in generalizing only to that population. Most of our experiments have been done on college students, and we know that such samples are not representative of the population at large. Indeed, the great majority of our experiments have been with elementary psychology students, and such samples may not even be representative of the college population of a school. Furthermore, because of differences in curricula and differences in entrance requirements, elementary psychology students at different universities do differ. In short, we are not on firm ground if we generalize from elementary psychology students to the population as a whole unless we have independent evidence that our given finding is not related to population differences. It must be understood that psychologists have not been unaware of the fact that the behavior of elementary psychology students is not representative of the eighth-grade pupils in Storm Lake, Iowa, or of the rat colony maintained by Harvard University, or of the employees of Armour and Company. The use of elementary psychology students has been an expedient.

It is not entirely unwarranted, of course, to assume that the laws of behavior derived from college students may have *some* generality. After all, college students may be defined as humans and many kinds of laws may be presumed to be general among humans. Be it noted, too, that it is well worth-while to have general principles of the behavior of college students. Sooner or later, we assume our experimentation will be extended so that we sample a much larger and heterogeneous population than heretofore has been available. In the meantime, our generalizations must be made with due caution.

The material or task. Task variables are usually important for most phenomena. What may hold true in learning a poem may not hold true in learning to fly an airplane. Thus, even in generalizing to a restricted population for a given phenomenon, E should first have explored representative types of tasks or materials with respect to this phenomenon. This, in most cases, is a step which can and is taken by most psychologists.

The specific conditions. Not only may a phenomenon be limited to a given kind of material for a given population, but it

may also be limited to a very restricted manner in presenting that material. For example, one phenomenon (reminiscence) appears when nonsense syllables are presented at a 2-second rate on a memory drum but not when they are presented at a 4-second rate. The continual occurrence of a relationship under several different conditions must be necessary before generalization to other conditions is accepted.

Dimensional extremities. Almost any scientific generalization concerning the influence of a given variable breaks down when that variable is manipulated at its extremes. Probably the reason for this is that other variables are automatically brought in when the main variable is effective in an extreme degree. The activity of a rat increases as deprivation time increases, but this obviously cannot be carried to its logical extreme; with extreme deprivation time the rat weakens. In the previous chapter we have said that learning-how-to-learn transfer seems to be uninfluenced by the passage of time. If the elementary psychology students on whom this principle was derived were tested on the second list when they are 80 years of age, we would probably be forced to amend the principle. Almost all psychological relationships between stimulus dimensions and response dimensions explicitly or implicitly carry the qualifying phrase, "within limits." The breakdown of the relationship at the extremes of the stimulus dimension is to be accepted.

These, then, are the four factors to bear in mind when consideration is being given to possible generalizations. All generalizations represent approximations to the truth, and statistically, approximations are appraisable by the error of estimate. One guide may be kept in mind when we are faced with the problem of scientific generalization: the further and further one's generalization gets beyond the conditions and subjects from which the principle was derived, the larger and larger will be the probable error in its wider application.

CHAPTER XI

Learning I — Conditioning

LEARNING: INTRODUCTION

> Learning so pervades human activity that any curiosity about the nature of man and his behavior leads sooner or later to inquiry about how his habits are formed, how his skills are acquired, how his preferences and tastes develop, how his knowledge is obtained and put to use. Equally important is how he becomes enslaved by prejudice and bigotry and other learnings which lead to trouble instead of to a satisfactory solution of his problems (Hilgard, 151, p. 1).

Definitional problems. We infer the characteristics of the learning process from observed performance changes. Learning is thus a construct: it is not something we observe directly. When we set out to specify the characteristics of the performance changes from which learning can be inferred several considerations should be kept in mind:

1. The performance change must result from *practice* or repetition of a specific behavior sequence. This requirement differentiates performance changes associated with learning from performance changes associated with maturation. Changes in performance indicating maturation are attributable to neuromuscular-glandular growth, and not to repetition of a specific behavior sequence.

2. The change must represent an increment in performance, where increment is demonstrated by response measures agreed upon by *Es*. This qualification differentiates learning from fatigue, since fatigue is defined as a decrement in performance. Nothing is implied concerning the moral aspects of performance. "Bad" habits as well as "good" habits represent increments in performance.

3. The stimulus manipulation which we relate to performance increment in order to define learning consists of at least two repetitions of a given sequence. Learning, being a *change* in performance, can be inferred with accuracy only if two or more measurements are taken.

4. In a very strict operational definition we might conceive of two matched groups, one group (experimental) being given one or more repetitions of a situation-response sequence, the other (control) being given none. If a test of the responses of both groups discloses better performance by the experimental group, we say that learning has taken place.

5. Keeping the above restrictions in mind, we may say that *learning is the acquisition of new responses or the enhanced execution of old ones.*

Chapter plans. Four chapters will be formally devoted to the study of learning. We say *formally* devoted because it is quite apparent that the preceding five chapters have often referred to the learning process. Even Chapter X, in which we studied experimental design methods, would not have been written were it not for the inevitable practice effects (learning) which accompany experimentation and which must be controlled adequately in an experimental design. We have seen how motivation is inextricably related to learning; we have seen how frustration experiments usually involve placing *S* in a learning situation and then making it difficult or impossible for him to learn. Transfer of training is a general variable which determines in part the rate at which learning will take place. Our study of learning proper will thus only focus attention on a highly important area with which we are already generally familiar.

The four chapters on learning are, in order, *Conditioning, Multiple-Response Learning, Thinking,* and *Theoretical Problems.* The expediency of that order is derived from the author's conviction that one should proceed from the simple to the complex. Relatively speaking, this order of the chapters advances in steps according to the level of complexity of the situation in which learning is studied. Yet there is nothing absolutely simple about even the least complex situation—which is conditioning—since as we shall see, a conditioned response is an extremely complex bit of behavior.

Another reason for considering these learning phenomena in the order indicated is that principles derived from conditioning are often used as explanatory devices for more complex behavior. Our use of the concept of *generalization* in studying conflict and transfer of training is an illustration of this procedure. The transfer of principles from one level of complexity to another has been largely uni-directional, i.e., principles of the conditioned responses have been used to explain more complex phenomena, but the principles derived at the complex level have seldom been used to explain conditioning phenomena. The tendency to apply conditioning principles as explanations of more complex behavior emphasizes the analytical nature of scientific procedure. Empirical relationships discovered at the simple level make good starting points for analyzing complex situations. This procedure does not, of course, mean that the laws of the simple case will inevitably hold in the complex one, but it does give the scientist a starting point both in method and toward the construction of theory.

CONDITIONING: INTRODUCTION

Conditioning as an experimental procedure and as a mine of behavioral data has become thoroughly established as a part of American psychology. In spite of the different points of view which prevail as to its particular place in psychology, there can be no doubt of the status of conditioning when evaluated by the number of psychologists whose research interests center around it. The most exhaustive work in the field (153) published in 1940, even though largely concerned with American investigations, lists 973 references. During the years since that publication, scores of new researches have been published. In our one chapter we shall be able to include only the high points of conditioning procedure and concepts.

Conditioning is defined by a cluster of techniques. The conditioned response obtained by a given technique is usually a relatively discrete response to a relatively discrete stimulus. We say relatively because the response, or aspects of it, may be measured in several different ways, and we say relatively again with regard to the stimulus because in some instances there is not a discrete stimulus event but rather, a stimulus complex. By and

large, however, in all conditioning we are able to point out a given response which occurs to a given stimulus after training, though before training was undertaken no such associative connection existed.

The plan of the chapter. The sequence of sections for this chapter is as follows:

1. Procedural variations defining conditioning
2. Methods of response measurement
3. Variables determining rate of conditioning
4. Variables determining extinction
5. Other phenomena
6. Concluding considerations

PROCEDURAL VARIATIONS DEFINING CONDITIONING

The extensive literature forbids here any attempt to classify the great variety of responses which have been conditioned. The student is referred to Hilgard and Marquis (153, Chapter II) for this information. We may mention only that nearly all types of animals on the phylogenetic scale from one-celled organisms through man have been successfully subjected to conditioning procedures. Our attention turns to the two most-used conditioning procedures. Under each of these procedures there are several specific techniques.

Procedure I

This procedure is exemplified by Pavlov's (284) work with the salivary response in the dog and includes all procedures commonly known as *classical conditioning*. We are already acquainted with the general procedure, so we need only make formal definitions of the aspects of the procedure:

Conditioned Stimulus (*CS*): Any stimulus which before training *will not* evoke the response to be conditioned but will evoke it after training.

Unconditioned Stimulus (*US*): Any stimulus which before training *will* evoke the response to be conditioned.

Unconditioned Response (*UR*): The response elicited by the *US*.

Conditioned Response (*CR*): The response which is elicited by the

CS as a consequence of the training procedure. This *CR* is the performance change from which we infer learning.

Pavlov discovered early in his researches that in order to develop a stable *CR* most rapidly it was necessary to eliminate all distracting stimuli. Consequently, most of his later work was done in sound-proofed rooms with all stimuli presented mechanically. Most American investigators have followed Pavlov's lead in this matter. Mechanical timing devices, careful adaptation of the organism to the experimental situation before training starts, the use of sound-proofed rooms, and so forth, have all been common in the experiments we shall review.

Let us diagram the training sequence of Procedure I. The dotted lines represent temporal duration of the stimuli, and Pavlov's usual stimuli are used as illustrations.

CS ....:
(bell)
US ....:
(meat powder) → *UR*
(salivation)

Except when the time interval is a variable of study the *CS* usually precedes and overlaps the *US*, since this relationship gives more rapid learning. The *CS* is followed by the *US* in an invariable sequence. The *US* produces the *UR*, salivation (in the above case). The pairing of the two stimuli continues time after time until the *CS* will by itself elicit the salivation. This response, the *CR*, need not be a replica of the *UR*, but is usually quite similar. Determination of whether a *CR* has been formed is accomplished in two ways:

1. If the response *anticipates* the *US*, it is an indication that it is occurring to the *CS*.

2. *E* may give *test trials* at various points in the training when only the *CS* is given. If the response occurs to the *CS* this is, of course, indicative of conditioning.

Distinctive features of Procedure I. Analysis of the Pavlovian-type procedure indicates that the method has four distinctive features, and we may use these features to distinguish Procedure I from Procedure II.

1. *US* produces the response to be conditioned. It will be noted that in Pavlov's procedure the *US*, meat powder, produced salivation, a response which eventually became associated with the *CS*.

2. *S* does not discover the "correct" response—it is "forced" upon him by the *US*. The presentation of *CS* and *US* is independent of *S*'s behavior.

3. *S* is usually physically restricted with but little freedom of movement allowed.

4. There is a discrete *CS*.

Illustrations of Procedure I. In addition to Pavlov's work there are several other specific techniques which may be classed under Procedure I and which have been used with considerable frequency.

1. *Flexion response in animals.* In several instances in the preceding chapters we have seen how *E* has conditioned a dog to lift his foot to the sound of a buzzer (*CS*) and thus avoid a shock (*US*). In some experiments the shock has been administered whether the flexion response is made or not; that is, even though the animal responds to the *CS* by lifting its foot.

2. *Finger withdrawal.* In conditioning finger withdrawal *E* may or may not allow the response to avoid the shock.

3. *Eyelid closure.* This has been a favorite response to condition in humans, and in view of the fact that we have not previously had an illustration of such conditioning, the details of the procedure should be indicated. Rather elaborate apparatus has been constructed for conditioning this response in order to assure precise stimulus control.

The *S* is usually seated in a dark room with his head held firmly in position. *E* clamps a small tube onto the apparatus holding the head in position so that its outlet is very close to the eye. At the appropriate time the tube will emit a slight jet of air directed at the eyeball, causing a reflex wink. The air jet is the *US*, the wink the *UR*. A slight hammer tap or a shock just under the eye has also been used as *US*s. A weak light or a "click" may be used as *CS*.

Measurement of the response is usually accomplished by photography. A false eyelid consisting of a very light but stiff strip of paper is pasted on the real lid and reflects a beam of light thrown directly in front of the face at right angles. When the

lid blinks, the false lid also moves and the shadow movement is recorded photographically. A timing system is so arranged that the events occurring between the presentation of the *CS* and the blink following the *US* are clearly outlined on the permanent record for each trial. As conditioning takes place the eyeblink will be seen to move ahead of the *US*, thus indicating that it is occurring in response to the *CS*. The method has also been successfully used with dogs (152).

4. *The psychogalvanic response.* This has been widely used in conditioning procedures with human *Ss*. The *US* is usually a shock, and as we know, the response is measured by a galvanometer as changes in electrical resistance.

Human voluntary responses are difficult to condition *in the laboratory.* This should be emphasized because it seems likely that many of our responses in everyday life are acquired under conditions which fit the conditioning paradigm. This is seen most clearly in young children as they learn the implications of verbal symbols. Parents often say "no" and follow this with a mild slap on the child's hand if he reaches for a breakable object. A few such occurrences usually result in the "no" itself eliciting withdrawal behavior, whereas in the beginning this stimulus had no effect. With adult *Ss* in the laboratory conditioning is a different problem since the voluntary control may inhibit the occurrence of the desired response.

In conditioning voluntary responses, such as finger or hand withdrawal, instructions are very important. *E* attempts, in a sense, to get *S* to dissociate himself from his hand and let the hand govern itself. *Ss* have sometimes reported that such a feeling of dissociation has developed. In some experiments *E* has required *S* to work simple problems while the finger was being conditioned. Here is a sample of instructions given by Wickens in his work on finger conditioning:

> This is a conditioning experiment. During the experiment your hand will be strapped to this board, with your middle finger resting on this electrode, and the other electrode fastened to your wrist. You will receive a shock which will pass from one electrode to another. You can break the shock by lifting your finger from the electrode. During the experiment a tone will precede the shock by a regular interval of time. If you get conditioned you will develop a tendency to respond to the tone before the

shock goes on. I do not want you to fight against becoming conditioned, and at the same time you are not to respond voluntarily to the tone. If your finger wants to fly up just don't inhibit it (408, p. 223).

To emphasize further the importance of instructions to *S*, let us consider an empirical study by Norris and Grant (281). These *Es* worked with eyelid conditioning to determine whether instructions *not* to blink would retard conditioning. The *CS* was a dim light, the *US* an air puff. Six conditions were employed but only two of these will be described here.

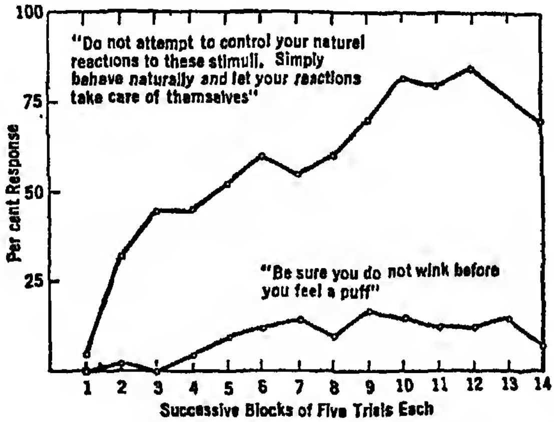


FIG. 56. The effect of two different instructions on eyelid conditioning. Data from Norris and Grant (281).

Before the initiation of the training trials, the control group was told: "Do not attempt to control voluntarily your natural reactions to these stimuli. Simply behave naturally and let your reactions take care of themselves" (281, p. 39). Members of the experimental group, on the other hand, were instructed: "Be sure you do not wink before you feel a puff" (281, p. 39). *Ss* in this latter group would, of course, wink to the puff, but the instructions were designed to inhibit the *CR* which after a certain number of training trials would normally begin to appear before the puff was administered.

Ss in both groups were given 70 training trials (*CS* followed by *US*), 40 of them being given one day and 30 the next. The mean percentage frequency of *CRs* for each block of 5 trials

for each group is shown in Fig. 56. The acquisition curves show clearly that instruction to inhibit the *CR* resulted in a marked reduction in frequency of *CR*s. Thus *S*'s set, as determined by *E*s' instructions and, we presume, by instructions *S* gives himself, may significantly influence conditioning of the eyelid reflex.

Procedure II

Procedure II is exemplified by the rat in the Skinner-box. A rat placed in the box must learn to press the lever to secure food. Let us note the differences in this procedure as compared with Procedure I.

1. The pellet of food does not produce the response which is to be conditioned. The response to the food may be chewing and swallowing, but the response conditioned is lever pressing.

2. *S* has to discover the correct response and the pellet is not given until this discovery is made. The rat may salivate, it may roll over, or it may do handsprings, but only when it presses the lever, can it obtain the food.

3. *S* is not as physically restricted as in Procedure I.

4. There is no discrete *CS*. The stimulus which elicits the response of lever pressing is better thought of as a stimulus complex in which the walls of the box, the lever, the placement in the box, the motivation, and so forth, all unite to form the stimulus complex in which the response is emitted. Basically, the diagram of Fig. 3, Chapter I, represents the paradigm of Procedure II as it is used in the present chapter.

Illustrations of Procedure II. In addition to the Skinner-box several other situations have been used with considerable frequency by various investigators.

1. *Problem boxes.* In the problem box *S* is to discover a response which opens a door and allows exit to obtain food. This is exemplified by Guthrie's work with cats (131). Cats are placed in a small box in which there is an upright pole. When the cat hits the pole a contact is broken causing the door to open and allowing the cat to reach a bit of salmon.

2. *Straight alley maze.* A straight alley maze is usually about 3 feet long. The animal is placed at one end and food at the other. The measure of learning is the increase in speed with which the animal runs from the starting box to the food box on

successive trials, or the latency in leaving the starting box on successive trials. This technique has been developed largely by Graham (115) and his students at Brown University.

3. *T-maze.* We are already familiar with this device. The rat must learn to take one or the other of the two arms in order to obtain food. Typically this is used as a discrimination problem in which the correct response is to take the arm in which a light is turned on and avoid the dark arm; to take the arm in which there is a square stimulus form and avoid the arm with the circular stimulus form; to take the left arm and avoid the right, and so forth. Frequency of correct responses on successive trials is the basic measure.

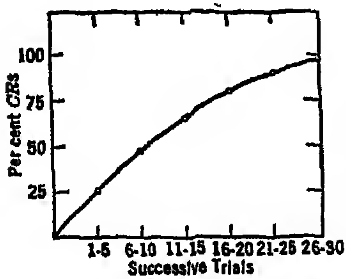
These two procedures, I and II, will not include all situations in which conditioning has taken place, but they will include the majority of them. So far as can be told at present regarding the factors which produce differences in the rate of conditioning, no great difference exists in responses learned in the two types of situations. Consequently, in most of our later discussion of the manipulable variables, we will make no distinction between Procedure I and Procedure II. The distinction does, however, have importance for the theoretical formulation to which we will turn in Chapter XIV.

MEASUREMENT OF CONDITIONED RESPONSES

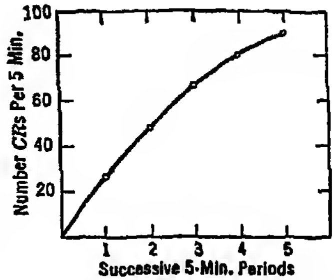
Several measures have been used as indices of the degree of conditioning attained during a series of trials. We shall name them and suggest how some of them are plotted in graphic form (Fig. 57). The curves are for illustrative purposes only, since there is no one *true* learning curve. All learning curves are presumably a function of the amount and character of the transfer of previous experience as well as a function of immediately pertinent conditions. Since many empirical curves are negatively accelerated in that each additional trial adds less and less to the measure of response strength, we have used this form in plotting the sample curves.

Frequency of the *CR*. This measure is very commonly used. Trials are plotted along the abscissa, with frequency of response along the ordinate. The frequency is usually in terms of

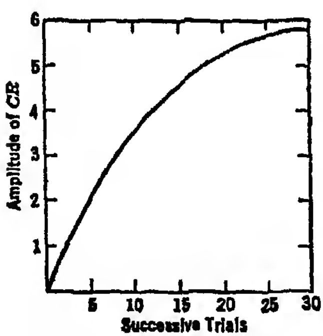
percentages, indicating the proportion of *Ss* showing *CRs* for a given trial, as determined by anticipatory responses or by test trials. If a considerable number of trials is given, certain trials may be combined, as are trials 1–5, 6–10, 11–15, etc., plotted in Fig. 57a.



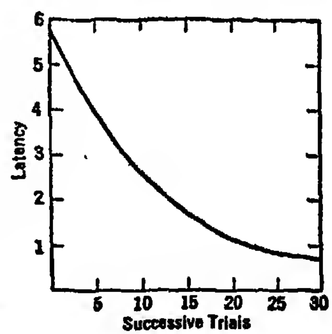
57a. Mean frequency of response per trial.



57b. Mean frequency of response for successive 5-min. periods.



57c. Amplitude of response.



57d. Latency of response.

FIG. 57. Frequently used methods of plotting the acquisition curve of conditioned responses. The values are hypothetical.

In Procedure II situations, such as those in Skinner-box learning, the frequency of response per unit of time is plotted. The greater the frequency the greater is the degree of conditioning. Such a curve is shown in Fig. 57b. The figure might also show days along the abscissa. Thus, if each rat were given 15 minutes per day in the box for 10 days, 10 units along the base line would represent the 10 days' training.

Amplitude of *CR*. This measure is commonly used in eyelid conditioning (amount of closure) and in psychogalvanic conditioning (amount of resistance change). It was also the measure most persistently used by Pavlov in measuring saliva secretion. Thus the increase in conditioning was shown by the increase in the number of drops of saliva (per unit of time) occurring to the *CS*. Plotting may be done as it is in Fig. 57c.

Latency of *CR*. This is another measure which is frequently used, in both Procedures I and II. The shorter the latency of the response the higher is the degree of conditioning. If a rat is taught to run down a straight alley, from a starting box at one end to the food box at the other, it is presumed that learning is shown by the decrease in latency between the opening of the starting box and the exit of the rat from that box. The learning curve in such an instance is shown as a decreasing function of successive trials (Fig. 57d).

Trials to reach a criterion. Data in these terms are often presented in tabular fashion. Thus *E* may decide that the arbitrary criterion of conditioning which he will use is 23 *CR*s out of 25 presentations of the *CS*. If two groups were treated differently, the variation in the speed of conditioning would be measured by the difference in the number of trials to reach the criterion.

These are the usual methods of reporting the quantitative aspects of the conditioning process.* Most *E*s agree that many responses are conditioned during the experiment but are not observed or not recorded. A rat may salivate when placed in a Skinner-box after having experience with food there, but the measure of conditioning is usually the number of times the lever is pressed, not the amount of saliva which flows.

Descriptive protocols (a running account of *E*'s observations) in conditioning experiments are sometimes available. Such observations may shed considerable light on the more strictly quantitative modifications of behavior which are recorded. Pavlov's work is rich in these descriptive protocols. Culler and his associates (74) have provided descriptions of the gross behavior modifications shown by a dog in flexion conditioning. Early in the training process the response to the shock is a violent struggle

*Resistance to extinction has sometimes been used as a measure of conditioning. Extinction is discussed in a later section of the chapter.

reaction of the "whole" dog. Gradually, as the response begins to antedate the shock, gross bodily movements extinguish along with the struggling. In the well-established avoidance response the dog lifts his foot from the grid at the presentation of the *CS* and very little involvement of the remainder of the body is apparent. In some excellent movies which have been made of this conditioning process, the final response of the dog suggests that he is, anthropomorphically speaking, utterly bored with the whole procedure, for he lazily lifts his paw at the sound of the *CS*.

VARIABLES DETERMINING RATE OF CONDITIONING

Because of the extensive number of variables which influence the rate of conditioning, we shall not be able to treat any of them exhaustively. Our purpose will be to give an understanding of the methods of manipulating a variable, and we shall usually do this by an experimental illustration. Most of the illustrations will be from American studies, although Pavlov may rightfully claim credit for the discovery of many of the variables. In examining these variables we shall introduce additional phenomena which have been discovered during work on the variable. Finally, we shall attempt to state the general empirical relationship between a given stimulus dimension and the rate of conditioning.

These variables we are about to take up determine the rate of conditioning as this is measured by one or more of the methods previously outlined. As used here, rate of conditioning means the amount of change for a constant number of trials. Thus, at the end of 15 trials under one condition, we may find that the *CS* elicits 60 per cent *CRs*, while under a different condition it elicits only 30 per cent. Rate of conditioning in the first is greater than in the second. Closely related to rate of conditioning is *level of conditioning* attained. Most conditioning curves ultimately reach an asymptote or limit beyond which further trials produce little improvement for a given set of conditions. The general relationship between rate and level is positive in that the faster the rate of conditioning the higher the level ultimately attained. Consequently, if we present data which show that the level of conditioning attained for a constant number of trials varied for different conditions, we may infer that the rate of conditioning for the

different conditions also varied, providing, of course, that the different conditions represent variations along one dimension only. Conversely, if we show curves for different conditions which have a different rate of change, we may conclude that the level of conditioning ultimately attained would be different irrespective of the number of trials given. It is clearly possible, however, that the curves for different conditions might well reach their asymptotes in the same number of trials. The distinguishing characteristic is that the levels attained are different; hence, rate of change to reach the levels is different for the conditions under observation.

To the variables which we shall take up we add, of course, Motivation (Chapter VI), and in so far as it can be applied, Transfer of Training (Chapter IX). These two general variables have already been discussed.

Time Between *CS* and *US*: Procedure I

This variable is limited to Procedure I situations since we do not have a discrete *CS* in Procedure II. It has been customary to give special names to various temporal relations between *CS* and *US* (284):

Simultaneous CR: *CS* precedes and overlaps *US*; time between onset of *CS* and *US* may vary from 0 to 5 seconds.

Delayed CR: *CS* precedes and overlaps *US*; time between onset of *CS* and *US* is 5 seconds or over.

Short Trace CR: *CS* precedes but does *not* overlap *US*; time between cessation of *CS* and onset of *US* is from .1 to 5 seconds.

Long-Trace CR: *CS* precedes but does not overlap *US*; time between cessation of *CS* and onset of *US* is 5 seconds or over.

Backward CR: *US* precedes *CS* and does not overlap it.

Because so little work has been done on delayed and long-trace *CRs*, we are not going to consider them except to say that they are very difficult to form and usually cannot be developed until after a simultaneous *CR* or short trace *CR* has been established first (316, 317, 374).

Simultaneous *CRs*. Several studies have shown that the most rapid rate of conditioning takes place where the *CS* precedes the *US* by a short period of time, varying from .2 seconds to 1 second,

depending upon the particular response being conditioned. Simultaneous onset of *US* and *CS* is not a favorable condition for rapid learning.

The method employed in studying this variable may be illustrated by Kimble's study (197). Kimble conditioned the eyelid reflex, using six different intervals between *CS* and *US*. The *CS* was the brightening of a patch of light directly in front of *S*; the *US* was a jet of air against the eyeball. The *CS* was on for 1.5 seconds. The inter-onset intervals (interval between onset of *CS* and onset of *US*) for the six conditions were .1, .2, .225, .25, .35, and .4 seconds. As can be seen, the *CS* overlaps the *US* for each condition. There were 13 *Ss* in the three short-interval groups and ten *Ss* in the three long-interval groups.

Each *S* was given 60 trials—60 pairings of *CS* and *US*—and records of *CRs* were made photographically. As noted earlier there are two ways by which *E* may determine when *CRs* occur: (1) by observing the responses which anticipate the *US*, and (2) by the use of test trials on which *US* is omitted. If the time between *CS* and *US* is short, as was true in some of Kimble's conditions, the test trial method is mandatory, since there would not be time for the *CR* to anticipate the *US*. Therefore, Kimble used the test-trial method, introducing a test after every tenth training trial, so that there were six such trials during conditioning.

One other aspect of Kimble's procedure should be discussed. Pavlov has demonstrated a form of conditioning called ***temporal conditioning***, in which no discrete *CS* was given, but food was presented to the animal at regular intervals, perhaps, at 1-minute intervals. After this was continued for a period of time, it was noted that the dog began to salivate just before the end of the 1-minute interval. Apparently, the temporal interval has acquired the potency of a *CS*, although it is likely that the true *CS* is some process(es) associated with the 1-minute interval. Because of this possibility of temporal conditioning, the interval between successive trials is varied unless *E* is studying the temporal interval as a CS. Kimble varied the interval from 1 to 2 minutes between successive trials so that little possibility of temporal conditioning existed.

The results of Kimble's experiment in terms of frequency

of *CRs* given on all test trials are shown in Fig. 58. We see that maximum conditioning took place when .4 second intervened between *CS* and *US*. Other results suggest that at .4 second Kimble had almost reached the optimum interval and that longer intervals, perhaps .5 second or over, would have shown a decrease in the curve.

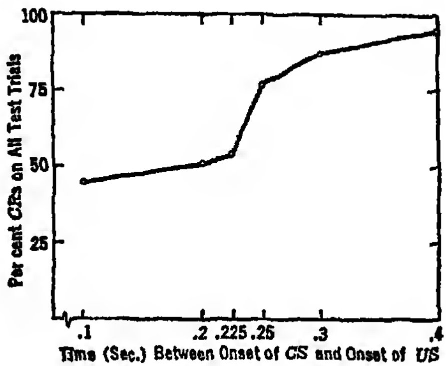


FIG. 58. Eyelid conditioning as a function of time between onset of *CS* and *US*. Data from Kimble (197).

Short trace *CRs*. In trace *CRs* the *CS* does not overlap the *US*. The procedure may be illustrated by Reynolds' study (309). The *CS*, a brief click delivered through earphones, was followed by a puff of air to elicit the eyelid reflex. Four intervals between cessation of *CS* and onset of *US* were used: .25, .45, 1.15, and 2.25 seconds. For each interval a separate group of *Ss* was given 90 trials distributed over 4 days. An average of 1.5 minutes intervened between each trial. Since the intervals were long enough to allow anticipation of *US* by *CR*, this anticipation was used to indicate conditioning. The four learning curves plotted in terms of per cent *CRs* for each block of ten trials are shown in Fig. 59. The greatest degree of conditioning took place with an interval of .45 seconds between *CS* and *US*, and the least conditioning with an interval of 2.25 seconds.

Let us compare the differences in the methods used to plot Reynolds' and Kimble's data (Fig. 58). Kimble's data are plotted in terms of total *CRs* on all test trials with the independent variable along the abscissa. From this graph we would infer that

the rate of conditioning was directly related to the total number of *CRs*, which in fact it was. On the other hand, the acquisition curves of Reynolds' experiment are plotted (Fig. 59), and differences in rate of learning are observable directly. Differences in level of learning attained as a function of the independent variable may be determined by comparing the frequency of *CRs* on the last ten trials (81–90) for the four conditions.

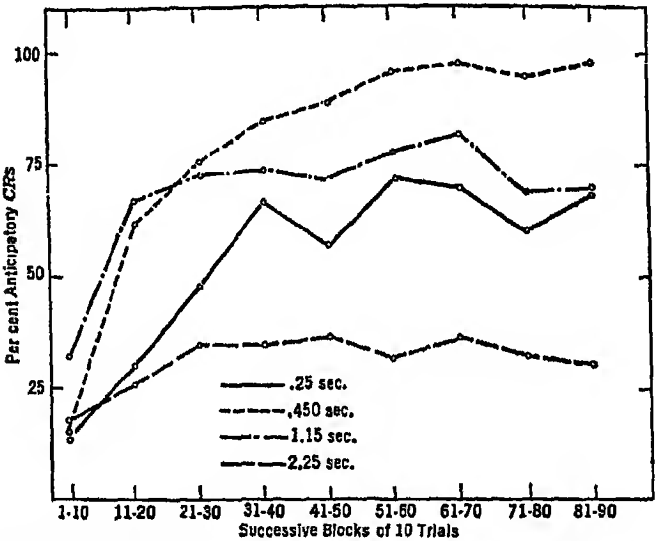


FIG. 59. Short trace *CRs* as a function of interval between *CS* and *US*. Data from Reynolds (309).

One additional matter concerned with eyelid conditioning should be briefly discussed. Grant (118) has recently discovered a non-learned eyelid response *to the CS* when the *CS* is a light. This response, which Grant calls the *beta response*, has a latency of .12 to .24 second. It will occur only when the intensity of the light used as *CS* is of a certain magnitude, and then only if *S* has been dark-adapted (119, 120). Some of the earlier studies on eyelid conditioning may have counted these beta responses as *CRs*. This possibility does not apply to Kimble's study because of the low intensity of the *CS*, or to Reynolds' study since he used a click as *CS*.

Backward conditioning. Backward conditioning is developed as a result of a training procedure by which the *CS follows* the *US*. A study (372) published in 1930 gave the first purported evidence of backward conditioning. Using a buzzer as a *CS* and a light hammer tap below the eye as *US*, *E* studied the eyelid closure. Schematically this appears as follows:

US . . . .
(Hammer Tap)
↘
UR . . . .
(Wink)
CS . . . .
(Buzzer)

During training the *US* always precedes the *CS*; hence, it was necessary to give test trials to determine if conditioning had taken place. After training, 18 out of 20 *S*s in this experiment gave evidence for conditioning, i.e., they responded to *CS* alone.

Some *E*s, using various responses and techniques, have been unable to find clear-cut evidence for backward conditioning (60, 296). Other *E*s have found only slight evidence for backward conditioning (e.g., 19, 284, 365, 419, 420). The fact that some *E*s have been unable to find backward conditioning, that others have found only slight conditioning, and that a recently discovered phenomenon may account for all backward conditioning as an experimental artifact, leads to the conclusion that *true backward conditioning has not been demonstrated unequivocally*. The phenomenon which may account for so-called backward conditioning is *pseudo-conditioning*.

Pseudo-conditioning. A study by Wickens and Wickens (409) will demonstrate the procedure which defines pseudo-conditioning. Infants under ten days of age were divided into three groups, and treated to the following procedures:

Group I: Buzzer (*CS*) followed by shock (*US*) to foot. The shock caused foot withdrawal. Twelve trials a day were given for three days.

Group II: Given 12 shocks a day for three days. *The buzzer was never sounded.*

Group III: Tested for response to buzzer before and after three-day period *but never shocked.*

At the end of the 3-day period, all groups were given tests in which the buzzer alone was sounded. The results show that *Ss* of Group II gave just as many responses (foot withdrawal) to the buzzer as did Group I *even though the buzzer and shock had never been paired.* That the shock is the important element in this finding is shown by the fact that Group III gave no responses on the test trials. Had these *Es* not used Group II they probably would have concluded, quite erroneously, that *CRs* were established in Group I.

Pseudo-conditioning, we see, is the application of *US* alone for several trials and then presentation of *CS* which has never been paired with *US.* If this *CS* elicits the response produced by the *US*, pseudo-conditioning has taken place. Recently, this phenomenon has been found by several investigators for a wide variety of responses in animals and men, although there is at present no general agreement on its interpretation. Some evidence points to generalization between *US* and *CS* as a possible explanation (410) but there are not enough data on this matter to permit a definite conclusion.

Pseudo-conditioning has two important implications:

1. The meager evidence which has been interpreted as proof of backward conditioning is very likely a demonstration of pseudo-conditioning which appears most readily when rather sharp or strong *USs* are used (such as shock). It is under these conditions that backward conditioning has been found. Until backward conditioning is demonstrated over and above pseudo-conditioning, we may conclude that evidence for backward conditioning is unconvincing.

2. Studies have shown clearly that genuine forward conditioning takes place on a far greater scale than could be accounted for by pseudo-conditioning (e.g., 142). Yet, many studies of forward conditioning in which no control for pseudo-conditioning has been used may have overestimated the amount of true conditioning which did take place. Until it is definitely shown that pseudo-conditioning is a function of associative factors such as exist in true conditioning, it may be advisable to use a control for pseudo-conditioning in carrying out experiments and set apart the "conditioning" obtained under these conditions from the conditioning obtained under standard procedure.

Time Between Response and Reinforcement: Procedure II

In Procedure II, which has no discrete *CS*, the only obviously pertinent time interval is that between execution of the response and receipt of reinforcement (reward). With animals, the data show that the longer the interval between response and reward the slower is the learning. Perin (286), for exemple, found that if 30 seconds intervened between pressing the lever in a Skinner-box and receiving the food pellet, no learning took place. Learning was markedly retarded if the interval was 10 seconds, and learning was most rapid if there was no delay between response and reward.

The principle that the longer the delay between response and reward the slower is the learning appeals to common sense. Very young children are likely to perform better for immediate rewards than for promised rewards in the distant future. Even as adults we do not like to contemplate too long an interval between effort and reward for the effort.

There can be no doubt concerning Perin's experimental results, but some question has been raised concerning their interpretation. We have seen how animals may get secondary reinforcement from a consistently present stimulus when a primary reward is received on successive trials (Chapter VI). If the animal is fed in a box several times, even though there has been a delay between a given response and eating, it is clearly possible that the food box may take on secondary reinforcing properties and obscure the true influence of the delay period between response and reward. That is, the animal may be immediately reinforced (secondarily) by merely entering the box in which it has been fed.

Perin recognized this possibility, and since he wanted to study the influence of primary rather than secondary reward, he modified the Skinner-box so that the rat pressed the bar, was delayed, and finally ate the food all in the same small compartment. It was under such conditions that learning failed to take place if the delay was 30 seconds between pressing the bar and getting the reward. Since previous studies had found that learning would take place if the delay was as long as 20 minutes, Perin's results suggest that these previous studies must have resulted from failure to control secondary reward possibilities.

Other investigators have doubted that Perin himself eliminated secondary reward; indeed, it seems possible that the secondary reward feature for all rats would be very constant—all rats were delayed in the same food box. Perkins (287) attempted to show how secondary reward value of goal boxes may influence learning in the delayed-reward situation. In a simple T-maze, half the rats were trained to make a right turn at the choice-point, and the other half, a left turn. The delay boxes—those in which the animals were held until an appropriate interval had elapsed—were placed just in front of the goal-boxes. For one group of animals this box was the same from trial to trial; for another group the box was interchanged with a second one so that each was used as the delay box half the time. The delay boxes had quite different interiors, which afforded the animal distinctly different cues in each. It is assumed that a delay box gains secondary reinforcing value as a result of being associated with the food which is received immediately upon release from the box. If the same delay box is used constantly, it is presumed that it will gain more secondary reinforcing power than would two distinctly different boxes which were regularly interchanged.

Perkins' results confirm the expectation. The control group which used the same delay box from trial to trial learned more rapidly and to a higher level than did a group for which the delay boxes were constantly changed; but both groups learned even though the delay was 45 seconds. Indeed, some learning was shown by another group in which the delay was 120 seconds even though the delay boxes were changed. Thus, Perkins' results, like Perin's, show that the longer the delay the slower is the learning, and that some of the learning very likely can be attributed to secondary reinforcing properties of delay boxes. However, if all delay-of-reward learning is due to secondary reinforcement, Perkins obviously did not eliminate all secondary reinforcing stimuli. He suggests that these other secondary reinforcing stimuli may actually be the proprioceptive stimuli which occur when the rat makes the correct turn in the T-maze. These proprioceptive stimuli themselves become associated with receiving food; hence, when the rat makes the correct turn, the stimuli resulting from the movement act as secondary reinforcing agents on the response under consideration.

Grice's experiment. Subsequent theoretical analysis by Spence (363) and empirical substantiation by Grice (124) tend to support the idea that proprioceptive stimulation produced by the rat's movements may become secondary reinforcing stimuli. A crude analogy to this hypothesized state of affairs may be found in the golfer who knows he has hit the ball well because of the "feel" of the swing and follow through. The "feel," having been associated previously with direct reinforcement of watching the flight of the ball, has in itself gained the capacity to reinforce ("nice shot").

Grice's conditions for testing the hypothesis were straightforward: he (1) devised one situation in which differential proprioceptive cues were at a minimum, and (2) created another situation in which the proprioceptive cues resulting from making the correct response were markedly different from those resulting from the incorrect response. The first was constructed by using a simple black-white discrimination situation with the position of the black and white stimulus cues alternated from trial to trial. Going-to-white was the correct response but since the white stimulus cues were sometimes to the left and sometimes to the right, no distinctive proprioceptive cues could be associated with going-to-white and getting food. The second situation was exactly the same except that making the response to the white stimulus cues involved going up an incline, and making the response to the black stimulus cues required the rat to weave through two narrow offset openings. The hypothesis suggested that these two distinctly different responses would provide proprioceptive cues to differentiate between the right and wrong responses.

As soon as a rat had made a response to the correct (or incorrect) stimulus cues, it entered the delay box. Delay time was varied, not by making the animal stay in a delay box, but rather by lengthening the alley between the point at which the response was made to the black or white cues and the goal box. Lengthening the alley simply forced the rat to take a longer time to reach the goal. Median times actually required to reach the goal box were used as delay times in handling the data.

Under these two conditions, which we find described in Grice's study, the delay time was 5 seconds. In the first situation, the one in which cues were minimized, the median number of trials to

learn to a criterion of 18 correct responses out of 20 trials was 580 trials. In the second situation, the situation in which different proprioceptive cues were provided, the median number of trials required to reach the same criterion was 295. Thus, the evidence indicates that proprioceptive stimuli may become associated with a food reward and thus serve as secondary reinforcing agents.

This series of studies not only emphasizes that greater delay of reward can take place because of secondary reinforcement, but also shows how important the forms of secondary or symbolic rewards are for learning even in the lower animals. The general principles to be particularly noted from this series are that (1) the longer the delay between response and reward the slower is the learning; and (2) the greater the elimination of secondary reinforcement cues, the more rapidly the learning rate falls off as a function of the delay interval.

TABLE 22

EYELID CONDITIONING AS A FUNCTION OF INTENSITY OF *CS*

Intensity of tone (decibels)	*Mean Frequency Anticipatory Responses*
15	24.3
30	25.5
45	25.7
60	23.6
75	23.5
90	30.1

Data from Carter (59).

Intensity of *CS*

It might be expected that rate of conditioning would be a function of intensity of *CS*. *This expectation is not substantiated by the data.* Carter (59), using six groups of *Ss*, has studied this variable in human eyelid conditioning. The *US* was a puff of air, and the *CS* a tone of 1024 cycles, the intensity of the tone being the independent variable. In terms of decibels above threshold, the intensity of *CS* for the six groups was 15, 30, 45, 60, 75, and 90. (120 decibels is approximately the threshold of pain.) After preliminary testing, 50 training trials were given and conditioning was measured by the mean frequency of responses which

anticipated the air puff. These data are shown in Table 22. There is no evidence of any consistent relationship. We may, therefore, conclude that according to the evidence available to date, *intensity of CS is not an effective variable.* If *CS* is above threshold, conditioning will take place.

Complexity of *CS*

Though the intensity of *CS* is not an effective variable, the complexity of *CS* definitely is important. By complexity of *CS* we mean the number of different components. The relationship will be illustrated by Miller's study (258).

He used the reflex eyelid response, the closure being effected by a *US* which was a slight electric shock below the eye. For these groups (ten *S*s to a group) the *CS*s were:

Group I: A black pointer moving on a white background and a simultaneous sound of changing pitch received through earphones. This is considered a compound *CS*.

Group II: Sound of changing pitch only.

Group III: Sight of moving pointer only.

Miller presented to each group of *S*s 50 paired reinforcements, the results of which are expressed by per cent frequency of anticipatory responses given by each group. Group I gave 81 per cent conditioning; Group II 45 per cent; and Group III, 23 per cent. The compound *CS* is much more efficient than either single *CS*, and the auditory *CS* produces more conditioning than does the visual *CS*.

Miller's data may be somewhat misleading. We see, for example, that level of conditioning attained with the compound *CS* is greater than that of the two single stimuli added together. The more probable relationship is that the level of conditioning produced by a compound *CS* is somewhat less than the arithmetic sums of the conditioning produced by the individual components (165). In Miller's experiment visual observation of the moving-pointer stimulus may have had some influence on the eyelid reflex, perhaps an interference effect, which was greater when the pointer was used as *CS* alone than when used in combination with the pitch stimulus. It is sufficient for us to know that under otherwise comparable conditions, compound or complex *CS*s produce more rapid learning than do unitary *CS*s.

Complexity of *US*

Although few studies have been concerned with this variable, enough data are available to show that it is an effective one. This is demonstrated in a study by Weber and Wendt (400), who investigated the eyelid reflex. Four groups of *S*s were used in a procedure in which the *CS* was always a light stimulus and the *US* was varied as follows:

Group I: US was very loud sound (approximately 95 decibels). (A loud sound will produce eyelid closure).

Group II: US was puff of air to eye.

Group III: US was puff of air *or* loud sound, these two being used randomly, but for the total period, equally often.

Group IV: US was *both* an air puff and a loud sound administered simultaneously.

TABLE 23

CONDITIONING AS A FUNCTION OF COMPLEXITY OF *US*

US	*Mean Frequency CRs in 50 Trials*	*Mean Amplitude (mm) of CRs in 50 Trials*
Sound	5.5	0.3
Puff	16.0	0.8
Sound *or* Puff	16.0	1.0
Sound *and* Puff	28.8	2.1

Data from Weber and Wendt (400).

S was given fifty training trials. The results (Table 23) show that the greatest conditioning took place with the compound *US*, the sound alone being a very ineffective *US*. The authors believe that the sound was ineffective because *S* rapidly adapts to it.

Pre-Association of *CS*s

In an experiment by Brogden (35) a *light* and a *bell* were presented together to dogs for 200 trials. Following this presentation a conditioned flexion response was established by using a bell as the *CS* and a shock as the *US*. On test trials the *light* was presented and it was shown that the flexion response occurred to the light, although *the light and shock had never been paired.* Adequate control groups were used in which the preliminary pairing of light and bell was omitted. The obtained phenomenon, called

sensory pre-conditioning, is a form of transfer of training which obviously influences the rate at which conditioning takes place. In two out of three attempts, sensory pre-conditioning has been demonstrated with human *S*s (37, 38, 187).

This phenomenon is an important one for learning theories. Somehow, during the pre-training presentations of light and bell, the two stimuli became associated or became functionally equivalent. Any theory which assumes that learning takes place only by primary reinforcement (food) or by secondary reinforcing stimuli, has some difficulty in pointing out the reinforcement which was present during the 200 pre-training presentations.

Massed vs. Distributed Training

In reading a description of this stimulus dimension one should think of massed training as the situation in which trials are given in rapid succession with a minimum time interval between each trial, and distributed training as the situation in which some longer interval is allowed between each trial. Experiments on this variable have been performed in three ways.

1. Number of trials is kept constant, but the interval between trials is varied. Thus *E* might give one group 50 training trials with 1 minute between trials, another group 2 minutes between trials, and a third, 4 minutes between trials. The exact times within any group, would have to be varied somewhat to prevent temporal conditioning. But, if there are differences in rate and degree of conditioning, they would be attributed to the major differences in distribution of the training trials. The work (number of trials)—rest ratio may, of course, be varied in many different ways.

2. A criterion is established and the number of trials to reach the criterion determined under massed and distributed trials.

3. The time interval is kept constant for a training period, but the number of trials given during the interval is varied. Thus *E* may use a 30 minute training period and give 15 trials spaced fairly evenly for one condition, and 30 trials during the same period for another condition. The results from such a procedure may not always be easy to evaluate, but in some cases where this has been used, it has been shown that a few distributed trials resulted in faster learning than did more massed trials (169, 309).

We need not take up a specific experiment. Several studies have shown that *conditioning takes place more rapidly by distributed than by massed trials.* In fact, in one experiment, a *CR* built up during distributed practice was eliminated if additional training trials were given under massed procedure (334).

Ratio of Reinforcement

The *US* in Procedure I and the reward (or escape from punishment) in Procedure II lead to reinforcement. The general impression which one would get from all that we have said thus far about conditioning is that reinforcement is necessary for strengthening responses elicited by a specific stimulus or by a stimulus complex. Later, we shall see that such an assumption is not shared by all psychologists. One of the reasons for this disagreement has been work on the variable—ratio of reinforcement. By ratio of reinforcement we mean the proportion of trials followed by reinforcement. In the usual experiment the reinforcement occurs after each trial. Much of Skinner's work (356) illustrated experiments in which the rat did not get a food pellet each time the lever was pressed; instead, the rat received a food pellet only at regular intervals regardless of how many responses were made. Skinner reports that one rat maintained a high level of performance with a food pellet received only after each 192 presses. However, the full importance of such results for theoretical formulations appears in a series of experiments by Humphreys (168, 170, 171). These experiments involved the conditioning of the psychogalvanic response, the eyelid reflex, and a lever-pressing response. Let us illustrate the finding with the conditioning of the psychogalvanic response.

One group of 20 *Ss* was given 16 paired stimulations of a tone (*CS*) and a shock (*US*). The *CS* was presented to another group 16 times, but only eight of these presentations were followed by the *US*, these 8 trials being distributed randomly among the 16. Thus, the latter group had only partial reinforcement—half as many reinforcements as the first group.

The results show that the conditioning at the end of 16 trials was approximately the same for both groups. In this case, it appears that a reinforcement does not necessarily strengthen the response; if so, why should the group with 16 reinforcements

not show a higher degree of learning than the group with only eight reinforcements? We might say that it was due to pseudo-conditioning, but other considerations suggest that the findings cannot be accounted for completely by psuedo-conditioning.

In studies using Procedure II (in which partial reinforcement has been shown to be as efficacious as complete reinforcement) it may well be that secondary reinforcement by symbolic rewards accounts for the findings. Denny (79) has shown that 50 per cent reinforcement is just as good as 100 per cent reinforcement in a T-maze, *providing* opportunities for secondary reinforcement are present during the responses which are nonreinforced by primary reward. If the source of secondary reinforcement is removed, 50 per cent reinforcement does not provide as rapid a rate of conditioning as does 100 per cent reinforcement.

To date, however, the results of partial reinforcement in responses conditioned by Procedure I remain unexplained by those (the writer included, 195) who hold that the *US* leads to a state of affairs which strengthens the response it evokes. A source of secondary reinforcement has not been identified in the Procedure I situations.

Summary

In this section we have been quite brief in the treatment of some of the variables influencing conditioning. Scores of supplementary studies have not been mentioned. Yet, we have covered the major determining factors which cause the rate of conditioning to vary, and, surely, by now the complexity of the conditioned response is evident. Furthermore, we appreciate something of the difficulty of experimental work in this area. Even control of the variables which have been identified is an exacting job, and it should always be kept in mind that some of the more complex problems have not been touched upon here. The extensive treatment by Hilgard and Marquis (153) should be consulted.

So that we may not leave this section with a feeling of confusion over the influence of so many variables, let us review them briefly and give tentative generalizations.

1. *Time between CS and US: Procedure I.* Optimum rate of conditioning occurs when *CS* precedes (and overlaps) *US* by a short interval, varying from .2 to 1.0 second for various responses.

2. *Time between Response and Reinforcement: Procedure II.* The sooner the primary reward is received after the response the faster is the learning.

3. *Intensity of CS:* This is not an effective variable.

4. *Complexity of CS:* The more complex the *CS* the more rapid is the conditioning.

5. *Complexity of US:* The more complex is the *US* the more rapid is the conditioning.

6. *Pre-Associations of CSs:* If one of two pre-associated stimuli is used as the *CS*, the other will provide faster conditioning than if it had not been associated with the first.

7. *Massed or Distributed Training:* The greater the distribution the faster is the learning.

8. *Ratio of Reinforcement:* Conditioning may take place as rapidly with proper partial primary reinforcement as with complete primary reinforcement.

FACTORS INFLUENCING EXTINCTION

Experimental extinction is a procedure introduced following conditioning. It is accomplished by presenting the CS on a series of trials without the US (Procedure I), or by removing the primary reinforcement (Procedure II). These procedures customarily result in the gradual decrement in the conditioned response as determined by the various indices of response strength. Following Pavlov, we say that this reduction is brought about by a form of *internal inhibition.* Internal inhibition is in contrast to *external inhibition* which is brought about by a *foreign* or *distracting* stimulus during training. Such a stimulus causes a *sudden, temporary* decrement in response to *CS*, whereas experimental extinction produces a gradual decrement in response. Extinction is not to be confused with forgetting. Extinction is an active experimental procedure, whereas forgetting of a *CR* is measured some time after conditioning without extinction intervening.

It is quite obvious that learning is an essential process if the organism is to adjust to changed environmental conditions. Not quite so obvious, perhaps, is the importance of the adjustive nature of extinction. The significance of extinction has been pointed out by Hull:

As a result of the largely random flux of events in the world to which organisms must react, it inevitably comes about that they will often be stimulated by extensive groups of conditioned stimulus elements, *none* of which is causally related to the critical factor or factors in the reinforcement situation. In such cases, if the stimuli evoke the reaction it will not be followed by reinforcement. This, of course, is wasteful of energy and therefore unadaptive. The necessarily unadaptive nature of an appreciable portion of the habits set up by virtue of the law of reinforcement naturally raises the question of how organisms are able to survive under such conditions. The answer is found in the behavioral principle known as *experimental extinction* (165, pp. 258-259).

This section treats briefly a series of factors which cause variation in the rate at which extinction takes place. Because the influence of some variables may be predicted by our knowledge of their influence on conditioning, and because the experimental method is clear, there will be no attempt to treat minor variables.

Number of Training Reinforcements

The inference is that the greater the number of reinforced training trials the stronger is the *CR*, although this is not without some complication as concerns partial reinforcement. However, if a response tendency is stronger as a consequence of greater number of reinforced repetitions, one would expect it to extinguish more slowly than would a weaker response. Indeed, we have seen (Chapter VI) that resistance to extinction is one index occasionally used for inferring differences in level of conditioning. The relationship between number of reinforcements and extinction rate is shown by Williams (414).

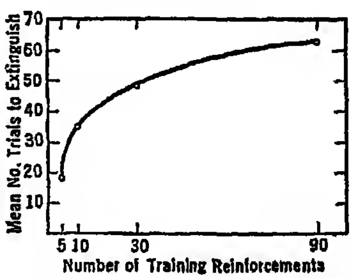


Fig. 60. Extinction as a function of number of training reinforcements. Data from Williams (414).

Williams used four groups of rats, 35 animals to a group. After premilinary training in the Skinner-box, Group I received 5 reinforced trials; Group II, 10; Group III, 30; and Group IV, 90. Each trial consisted of the rat's pressing the lever and receiving a pellet of food. In terms of

performance (time) on the first five trials, Williams could show that his four groups were roughly equal.

After 24 hours, extinction took place with all groups under the same deprivation state. During extinction the rats receive no food for lever-pressing, and the arbitrary criterion of extinction used was no-response-during-a-5-minute-period. The results (Fig. 60) clearly indicate that in terms of number of trials to reach extinction there is a direct relationship with number of training reinforcements.

Complexity of *CS*

The relationship between rate of extinction and complexity of *CS* is shown in a study by Razran (301), whose technique for studying conditioning is unusual. It will be remembered that Pavlov used the salivary response as the basic response for his studies with dogs. Razran studied conditioned salivary responses in humans. Dental rolls made of cotton were inserted under the tongue to absorb saliva. A roll was weighed (in centigrams) before it was placed in the mouth and then weighed again when it was removed, the difference in the two measurements giving an index of saliva flow. At least one conditioning trial was given for each roll. We know that without conditioning there would be some increment in the weight of the dental roll. Therefore control periods were necessary to determine the normal saliva flow. Flow above this in amount (as determined by weight) was considered to be a function of conditioning.

For *US*s Razran used various kinds of food. The *CS*s were visual in nature, so Razran misled his *S*s by telling them that the experiment had to do with the influence of visual fatigue on digestion. Conditioning was customarily obtained without difficulty.

In the experiment on complexity of *CS* and extinction, Razran used a great variety of light combinations for *CS*s. For extinction with a single *CS*, one of the units of the compound stimuli was used. In each case, of course, *S* was conditioned with the same *CS* used during extinction. In all instances he found that extinction was slower when the *CS* was complex. This is true even when adjustments are made for differences in acquisition level before extinction. Other studies have shown the same result.

We may conclude that extinction will be slower with a compound *CS* than with a unitary *CS*.

Massing vs. Distributed Extinction Trials

We have seen that distributed training trials enhance speed of conditioning as compared with massed trials. Now we ask how the massing and distribution of extinction trials influence rate of extinction. The data are not conclusive on this matter, as some studies (with different kinds of *CR*s) show little difference between rate of extinction as a function of massing and distribution, and others show some evidence that massing of trials facilitates extinction. There is no evidence that distributed extinction leads to faster extinction than does massed. Let us consider one study which in general shows that massed practice facilitates extinction.

Rohrer (320) used a device in which a rat was trained to flip a bar by an upward motion of the head. When the bar was flipped, a pellet of food was released. However, the rat was not free but confined tightly in a small wire cage and stock with only its head protruding. Furthermore, the cage was on a platform which could be moved away from the bar, and the bar in turn covered with a panel. By this method *E* could give discrete trials simply by moving the rat into position, recording the latency of bar-response, and then removing the rat to await another trial. This method fits neither Procedure I nor Procedure II as we have defined them.

By using several groups of rats, Rohrer manipulated not only the massing and distribution of extinction trials, but also the number of training reinforcements and the number of hours of deprivation at the time of extinction. The latter two variables become important in interpreting the result of massing and distribution, but the specific conditions are not pertinent to our momentary problem. Massing of extinction trials was defined by conditions in which 10 seconds were allowed between trials, and distributed extinction by 90 seconds between trials. Extinction was arbitrarily said to be complete when the animal failed to hit the bar in 10 seconds after exposure to it on two successive trials.

The results show that with low strength of original habit (as defined by number of reinforcements) there was little difference in massing and distribution of trials, whereas with strong initial

response strength, massing facilitated extinction.* As we would expect, the groups with the strongest initial habit took the longest time to extinguish. The groups with longer deprivation periods at time of extinction also took longer to extinguish. In general, the results support the general conclusion that when conditions are such (strong response strength and strong motivation) that a relatively large number of trials may be expected before a response is extinguished, massing will hasten the extinction. This empirically derived generalization leads to a consideration of the nature of the extinction process.

A bit of theory. Because we are to devote a special chapter to learning theory, we have avoided theoretical problems in considering experiments in so far as that has been possible. However, theory has so tended to dictate the design of extinction experiments that it is difficult to give a rationale for the experiment without considering theory. Briefly, let us look at one point of view for which there is some evidence.

Let us assume that when a stimulus evokes a response, an inhibitory state develops. If the response is reinforced, an excitatory state is built up. Let us assume further that the excitatory state is such that when the stimulus is presented again, the response will tend to be elicited. This means that the inhibitory state which is developed is less inhibitory than the excitatory state is excitatory. Or we might say that the unit of inhibition for a single response is -1, while the unit of excitation is $+2$, thus leaving a net effect of 1 in favor of the excitatory state. On successive trials the excitatory state gains more and more over the inhibitory state and we say learning is taking place.†

Now, let us assume further that during a period of rest, both the inhibitory and excitatory processes dissipate—grow less—but that the dissipation of the inhibitory state is more rapid than the

* Rohrer's groups of animals were obtained by random selection. His results provide one of the few illustrations available wherein random selection did not result in equal groups of rats. Two of his groups ($N = 20$) which were treated exactly alike during learning were shown to have significantly different mean latency in bar-hitting at the end of training. He was able to make certain adjustments in his groups by re-grouping so that the results are probably fair estimates of the true influence of experimental conditions.

† This is a much-simplified version of theories which have been suggested by various writers. The most complete form of such theory is to be found in Hull (165).

dissipation of the excitatory state. If such is true, rest periods during learning should result in more rapid learning. This would be expected because the rapid disappearance of the inhibitory state would leave the positive or excitatory state relatively stronger. We have seen that distributed practice results in more rapid acquistion of *CRs*; hence, the theory squares well with those data.

Now, let us consider what happens during extinction. During extinction we would have no new excitatory increments—only inhibitory increments. Consequently, over a series of trials we would expect the inhibitory state to become more powerful than the excitatory, resulting, in the end, in the elimination of the *CR*—experimental extinction. If we give distributed practice during extinction we would predict less rapid decrement than if we give massed. This would be true because during the rest interval the inhibitory state would dissipate. Under massed trials it would not dissipate; hence, extinction should take place more rapidly. We have seen that the data do not negate the theory as applied to extinction, but neither are they entirely favorable. Massing of trials does not always result in faster extinction.

What might be identified with the inhibition which develops during extinction? It has been suggested (262, 275) that when an organism makes a response and is not rewarded, it is, in a sense, punished. On subsequent occasions it should have less and less tendency to make that response. We may conceive of this as an avoidance response which grows stronger with each non-reinforced extinction response until it is dominant over the approach response (excitatory tendency) built up during the learning trials. At this point, extinction should be complete, as this is the point where the avoidance response is stronger than the approach response.

Furthermore, on the basis of such a theory, we might expect that anything which increases the punishment associated with failure to receive reward would enhance the speed of extinction. Mowrer (275) has guessed that if an organism has to work hard in making a response, the avoidance tendency built up will be greater than if the work were relatively easy. It is as though a rat says: "I did all that work and what do I get" in the hard work situation, and in the easy: "Well, I'm not out much, even though

I didn't get anything." Fortunately, even though one can state it in such flagrantly anthropomorphic form, Mowrer's hypothesis could be tested empirically. The idea is very straightforward: make the animal work hard in one case and not in another. If the hypothesis is tenable, extinction will take place more rapidly in the first instance. This bit of theory leads to the next variable.

Extinction as a Function of Amount of Work

Mowrer and Jones (275) trained rats in a lever-pressing device on which various size weights could be fixed. The heavier the weight the more difficult it was to push the lever and secure the pellet of food. Rats were trained to execute the response equally well for large and small weights so that at the end of training, the animals would respond just as rapidly with an 80-gram weight as with a 5-gram weight. This was a necessary step if different groups of animals were to be extinguished with different weights. Thus, all animals went into the extinction period having performed at an equal level during training with various size weights requiring different amounts of work.

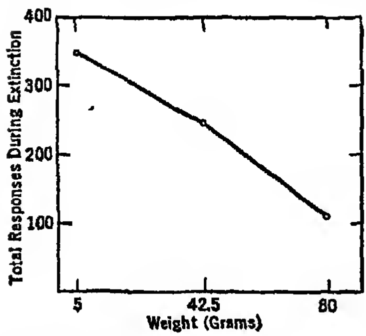


FIG. 61. Extinction as a function of the amount of work. Data from Mowrer and Jones (275).

For extinction, the 30 rats were divided into three groups of 10 each. One group went through extinction pushing an 80-gram weight, a second, pushing a 42.5-gram weight, and a third, a 5-gram weight. For three successive days the animal was placed in the box for 20 minutes, the measure of extinction being the mean number of responses made under each of the three conditions.

The results (Fig. 61) clearly support the hypothesis: the greater the size of the weight, the fewer are the responses made. It appears that a competing response—an avoidance response which competes with the pressing response—is established more rapidly the greater is the effort involved in making the lever-

pressing response. These results have also been confirmed by Solomon (360) in a jumping stand procedure wherein effort was manipulated by varying the distance the animal had to jump.

The Effect of Punishment

In the Mowrer-Jones study discussed above it was inferred that the labor involved in pushing the lever resulted in a form of punishment. In a sense, the punishment, as such, is of internal origin. We may ask what the effects of punishment will be if that punishment is supplied externally by *E* in the form of electric shock. Will extinction be hastened?

The answer to this question is given in a series of studies by Estes (91), who used a Skinner-box situation and introduced shock of varying intensity and duration following different levels of learning. The general results of this work indicate:

1. Punishment will immediately reduce the frequency of lever pressing, and as long as shock is continued the response will remain at a low frequency level.
2. If punishment is given early in extinction, the rate of responding will be suppressed immediately, but if the shock is subsequently removed, the rate of responding will rapidly increase and be followed by the decrease expected as a result of non-reinforcement.

In short, Estes' work gives little evidence that punishment in the form of electric shock at the time of response hastens extinction. The influence of punishment provided by shock is not the same as the punishment provided by hard labor in the Mowrer-Jones experiment.

Summary

There are minor factors which to some extent determine the rate of extinction and the shape of the extinction curves (153). However, we have covered the major variables and methods used to study them. Undoubtedly, there are variables which have not been discovered. Let us summarize those which were reviewed:

1. *Number of training reinforcements.* The greater the number of reinforcements during acquisition the slower is the extinction.
2. *Complexity of CS.* The more complex the *CS* the slower is the extinction.

3. *Massing vs. Distribution.* The evidence is not clear-cut, but it tends to support the generalization that extinction is most rapid with massed extinction trials.

4. *Amount of work.* The greater the work required of the organism in making the response, the faster is the extinction.

5. *Effect of Punishment.* There is no evidence that total trials to extinction is significantly reduced by punishment.

As yet, little direct attention has been paid to the rôle of secondary reinforcement by symbolic reward in extinction. Any *CS* or *CS* complex takes on secondary reinforcing properties so that extinction actually consists of removal of reinforcing properties of symbolic rewards. The more we know about the rôle of symbolic reinforcement the more we will know about extinction.

OTHER PHENOMENA

Spontaneous Recovery and Disinhibition

If a period of time is allowed to elapse following extinction and the *CS* is presented again, the response previously extinguished will have recovered a certain amount of response strength. This recovery defines the phenomenon of *spontaneous recovery.* Few American investigators have found that the response will recover to its former strength, though possibly not enough time has been allowed to elapse.

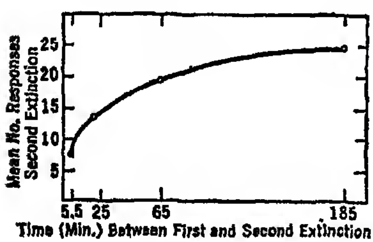


FIG. 62. Spontaneous recovery as a function of time following extinction. Data from Ellson (89).

Let us illustrate the general relationship between spontaneous recovery and the passage of time as determined by Ellson (89). Each of four groups of rats used in a Skinner-box situation was given 30 reinforcements and then extinguished. Spontaneous recovery was measured after various intervals of time in terms of number of trials to extinguish the second time, the criterion of extinction being a period of five minutes without a response. The first extinction required an average of 50 trials. The second extinction took place after 5.5 minutes

for one group, 25 minutes for another, 65 minutes for a third, and 185 minutes for the fourth. If a response recovers as a function of time, the longer the interval the greater should be the number of trials required for the second extinction. The results show that this is true (Fig. 62).

Spontaneous recovery and successive extinctions. Following spontaneous recovery *E* may again extinguish a response, once more allow recovery, extinguish again, and so forth. It is believed that only by such successive extinctions can a response be completely eliminated from an organism's repertoire of responses, though this has not been established. Some data have shown that each successive extinction takes place more rapidly as a function of previous extinctions, but we do not know whether it is due to the fact that extinction takes place more rapidly or whether it is because spontaneous recovery is less and less with each successive extinction. It is possible that if a response were allowed to recover completely after each extinction, all successive extinctions would take the same length of time. As yet we do not know the recovery properties of a response as a function of the number of extinctions, although it is highly probable that there is less and less recovery following successive extinctions.

Disinhibition. A phenomenon which must be related to spontaneous recovery is disinhibition. The operation necessary to demonstrate disinhibition is the presentation of a foreign stimulus, e.g., a raucous buzzer, during or immediately after extinction. Following this foreign stimulus, the *CS* is presented. If the supposedly extinguished response occurs at greater strength than was manifested just before the presentation of the foreign stimulus, disinhibition has been demonstrated.

In one study (373) a loud buzzer was sounded immediately following near-complete extinction of a psychogalvanic response. Then the *CS* was presented; the *CR* recovered nearly two-thirds of its former strength. The enhancement of response is short-lived, lasting only a few trials if the *CS* continues to be presented without the *US*.

Both spontaneous recovery and disinhibition represent a facilitation of the *CR* following extinction. Much less work has been performed on these phenomena than on acquisition and extinction. The most generally accepted theoretical interpretation of

spontaneous recovery posits an inhibitory process (as we have previously discussed) built up during learning which dissipates with the passage of time, thus leaving the excitatory state relatively stronger. The same inhibition is presumed to be "knocked out" in some fashion by the foreign stimulus which produces disinhibition; the foreign stimulus inhibits the inhibition.

Higher-Order Conditioning

In higher-order conditioning a *CS* serves functionally as the *US* for the conditioning of the *CR* to a new *CS*. Diagrammatically this procedure may be shown as follows:

First-Order *CR:* Bell (*CS*) ....:
Meat Powder
(*US*) ....
UR (salivation)

After several pairings the bell will evoke salivation, and a simple or first-order *CR* is said to have been formed. Then the following pairing is used to develop the second-order *CR:*

Buzzer (*CS*)..................:
Bell
(*US*) ....
UR (salivation)

The bell, by virtue of its previous association with salivation, will evoke the salivary response in the second-order *CR* so that after several pairings the buzzer will evoke the *CR*.

Occasionally, in establishing these higher-order *CR*s it is necessary to provide additional primary reinforcement. This would be expected, since it is clear that in higher-order conditioning we are dealing with secondary reinforcement through symbolic rewards, and the secondary reinforcing properties themselves may be extinguished gradually.

American investigators have been able to establish as many as five orders of *CR*s. Culler and his associates have done much of the work. An experiment by Brogden and Culler (39) is representative. Using foreleg flexion in dogs, they established the first order *CR* to a tone as *CS*. The *US* was a shock to the paw, but

after the response was established the shock was transferred to the thorax. It has been found that once the response has been established by shock to the paw, it may be perpetuated by a shock on other portions of the body. This is sometimes called *heterogeneous reinforcement* (153).

To establish the second-order *CR* a light was used as the new *CS* and the tone used to elicit leg flexion. Two additional orders were established so that the four orders were:

	CS	*US*
First order	tone	shock
Second order	light	tone
Third order	bell	light
Fourth order	electric fan	bell

The *Es* occasionally found it necessary to go back to primary reinforcement by administering the shock to the thorax following the use of the tone as *CS*, i.e., they went back to first-order conditioning. These *Es* found it progressively easier to establish the response as the order became higher, although it may well be that the animal reached a point where it responded to everything. Twenty-five trials were given each day, and the response was said to be learned when *S* made 23 responses out of 25 presentations of *CS*. The results (Table 24) clearly show that each successive *CR* was established more readily than the order preceding.

TABLE 24

TRAINING TIME FOR HIGHER-ORDER *CRs*

Animals	*Days*			
	1st order	2nd order	3rd order	4th order
A	28	18	5	3
B	25	19	6	3
C	46	22	4	3

Data from Brogden and Culler (39).

Mediated Conditioning

Mediated conditioning may be considered a form of transfer of training. The phenomenon will be illustrated by Graham's study (116). Hindleg flexion was established to a buzzer as *CS*,

and the inevitable shock was used for the *US*. Then, the shock to the hindleg was used as *CS* in order to establish *foreleg* flexion, with the shock administered to the foreleg as *US*. On test trials the buzzer was sounded. All four dogs used responded to the buzzer by flexing the foreleg, although the buzzer had never been used in direct association with foreleg flexion. Somehow, the conditioning of foreleg flexion to hindleg shock (after hindleg flexion had been established as a *CR* to the buzzer) served as a medium for bringing about the foreleg flexion to buzzer. Control animals were used so that pseudo-conditioning was ruled out, and it could be shown that either stage of conditioning alone was not enough to produce the results obtained with the experimental animals.

Mediated conditioning has some similarity to sensory preconditioning discussed earlier in the chapter. No explanations of these phenomena have been generally accepted; both pose difficult problems for learning theory.

CONCLUDING CONSIDERATIONS

Even in the brief survey of experimental procedures and results which has been presented here, few students can fail to grasp the complexity of the conditions which influence the allegedly simple conditioned response. It is certainly a gross euphemism to speak of the conditioned response as a simple, elementaristic unit of behavior. Yet, the various points of view concerning the "true" position which the conditioned response holds in the field of psychology are varied enough to make this systematic problem one which will plague any student who takes the matter seriously. We need to point out a few general guides to aid his thinking.

One error often made is to think of the conditioned response as a phenomenon totally divorced from other learning phenomena. This is, no doubt, to some extent caused by textbook writers who put the "conditioned response" in one chapter and "learning" in another. The modification of behavior shown by a salivating dog satisfies all the requirements of learning just as does the modification of behavior shown by a college sophomore when he finally recognizes the use of a standard deviation. We should

speak of "learning," not "conditioning and learning." In terms of the basic behavioral observations, "simple" conditioning belongs on the same dimension with the more "complex" learning which we will consider in the next chapters.

Students sometimes fall into the error of thinking that because most conditioning studies have been performed on animals or on involuntary and semi-involuntary responses in humans, the phenomenon must be very inconsequential for human behavior. Such a conclusion cannot be drawn—at least not on such a premise. "Real" emotions are difficult to produce in the laboratory; yet, few will deny their reality and importance in the lives of most of us. The fact seems to be that we don't know how much of our mature, "important" behavior has been formed under conditions which are counterparts of the conditioning paradigms. A little thinking on this matter will provide an almost endless list of possibilities. It is difficult, of course, to demonstrate that Edison's inventions came about because of a certain set of conditioned responses he had formed or that Shakespeare was able to produce *Hamlet* and *The Tempest* because of favorable conditioned responses which his environment produced. It is also difficult to produce Shakespeares and Edisons in the laboratory. We need take neither the extreme stand that maintains human adult behavior is a mere concatenation of conditioned responses; nor the other extreme point of view which holds that conditioned responses have no place in life outside the laboratory.

Experimental psychology must be analytical if effective variables and the function and interrelation of those variables are to be discovered. These analyses may be made at different levels of situational complexity. The conditioned response has been studied in what is believed to be the near lower limit of this situational complexity, namely, the analyses of the modification of simple responses to simple environmental changes. This is true only as long as we recognize the relativity of the term "simple" since, as we have pointed out, there is nothing absolutely simple about the conditioned response situation. It only appears to be simple when compared with situations in which a person memorizes the wing span of ten enemy planes, a situation which in turn appears to be simple compared with one in which a person learns to fly an airplane, which in turn appears to be simple as compared with

a situation in which a person designs a new airplane engine, and so forth.

In more advanced courses and reading the student will find that some psychologists believe that most of the basic laws of behavior modification can be deduced from the principles of conditioning. He will discover that some psychologists use a few basic principles of the conditioned response in explaining abnormal as well as normal behavior at all levels of complexity. He will discover that there are those who eschew the study of the conditioned response as being no more than an interesting laboratory phenomenon, largely without significance for adult activity. There are honest differences of opinion concerning the place of conditioning in psychology. These differences, in the writer's opinion, have come about largely because of basic differences in orientation or belief as to the level of situational complexity with which psychological study should begin. The student will find it almost imperative to suspend judgment of these points of view until an elaboration of them is possible in advanced courses.

From the material covered in this chapter the student should gain an elementary grasp of the experimental procedures used in studying conditioned responses and an ability to recognize that experimental finesse is a prerequisite to satisfactory science in even a "simple" situation. Regardless of the degree of complexity of the subject matter, the basic code of experimental conduct does not change.

SUMMARY

1. Conditioning is a group of techniques used to study learning. In the typical case, conditioning involves the acquisition of a relatively discrete response to a relatively discrete stimulus.

2. We have considered two general classes of procedure which are distinguished by (*a*) the discreteness of the conditioned stimulus, (*b*) the source of the response, (*c*) and the degree of physical freedom.

3. Four measures of conditioning have been commonly used: (*a*) latency, (*b*) frequency, (*c*) amplitude, (*d*) trials to attain a given level.

4. Eight major variables concerning the rate and degree of conditioning have been illustrated: (*a*) time between *US* and *CS*;

(*b*) time between response and reward; (*c*) intensity of *CS;* (*d*) complexity of *CS;* (*e*) complexity of *US;* (*f*) pre-associations of *CSs;* (*g*) massed vs. distributed training; (*h*) ratio of reinforcement.

5. Five major variables concerned with rate of extinction have been illustrated: (*a*) number of training reinforcements; (*b*) complexity of *CS;* (*c*) massing vs. distribution of trials; (*d*) amount of work; (*e*) punishment.

6. Other phenomena briefly discussed were: (*a*) spontaneous recovery; (*b*) disinhibition; (*c*) higher-order conditioning; (*d*) mediate conditioning.

In terms of a dimension of situational complexity, conditioning procedures define the least complex end of the dimension. In the next chapters we move up the dimension to the more complex situations in which learning is studied.

CHAPTER XII

Learning II—Multiple-Response Learning

INTRODUCTION

The previous chapter was largely concerned with unitary response associations—a single response associated with a more or less discrete stimulus. We considered methods of studying variables which influence the speed of the lever-pressing response, latency in leaving a starting box, frequency of lifting the forepaw at the sound of a buzzer, amplitude of eyelid closure to a tone, and so forth. The current chapter proceeds to a study of learning as it takes place in a more complex situation: namely, the situation in which a series of responses is combined or integrated into a patterned behavior sequence. In many cases it will be seen that this pattern consists of *an integration of a series of discrete responses to discrete stimuli.*

We are well aware that classifications made for purposes of study are not rigid. Perhaps the lever-pressing of the rat involves more than a simple "press"; almost certainly escape from a problem box constitutes behavior which can be called patterned behavior; the flexion response in the dog is, without question, not a simple muscle twitch. Yet, by and large, as compared with conditioning, we move up the dimension of task complexity from the less complex to the more complex when we study the integration of a series of responses.

A dimension of complexity may be related loosely to difficulty of learning. By most standards of evaluating difficulty, we would say that memorizing a passage of poetry is more difficult than acquiring a conditioned psychogalvanic response. Learning poetry requires voluntary effort on *S*'s part which is in contrast to what he does under the human conditioning procedure. Conditioned

responses of a primitive nature have been established in animals in which the higher brain centers have previously been severed from the lower centers which still function, and we have what is called *spinal conditioning*. It is difficult to conceive of a "spinal" John Doe memorizing a passage from Shakespeare. In a limited sense, then, complexity of task is related to difficulty of learning.

Rote learning. We often find the label *rote learning* applied to paired-associate or serial-type learning. The label implies a "dehumanized," and perhaps low-level form of learning. However, for mastering many tasks, it appears to be fairly efficient. New symbol series, foreign language and technical vocabularies, codes, and so on, can be rapidly acquired by rote learning. Although we might wish to think otherwise, this form of learning, to which this chapter is in part devoted, seems to be an essential core of the skills required in our schools. The mastery of many symbol systems by this method allows these symbols to become efficient aids to thinking and reasoning.

Plan of the chapter. The outline of the current chapter is similar to the preceding ones:

1. Tasks and procedures defining multiple-response learning
2. Methods of response measurement
3. Variables influencing rate of multiple-response learning
4. Special phenomena of serial learning
5. Maze learning
6. Concluding considerations

TASKS AND PROCEDURES DEFINING MULTIPLE-RESPONSE LEARNING

Within the scope of this chapter we will consider learning tasks which may be roughly distinguished from each other on the basis of the amount of overt activity required by the task. Thus, the memorization of a poem involves less overt motor response than does tracing a maze or kicking a football. The learning of verbal associations is considered to be near the lower end of a dimension defined as degree of overt response involvement, whereas tasks with a great deal of motor involvement are placed toward the other end of the dimension. There are many tasks which one would place between these extremes. Such tasks as

maze learning, typing, and code learning, show less motor involvement than does, let us say, swimming, but they show more than does verbal learning. We shall use this dimension as a basis for reviewing systematically materials and tasks which have been used rather frequently in experiments. Let us first discuss verbal tasks as representative of the lower end of the dimension.

Verbal Tasks

Any sort of written or spoken symbol may be employed in a learning experiment. It has become customary, however, to use certain classes of materials, and this custom has allowed some continuity from experiment to experiment, a state of affairs which is highly desirable for the integration of experimental results (253). We shall list the two major classes of material separately.

1. **Nonsense syllables.** We are already familiar with this form of material (Chapter IX). The original use of nonsense syllables was premised on the belief that they allowed learning to start at a point closer to zero than did more meaningful material. Another belief was that the variability of performance among *S*s would be reduced by using nonsense-syllables. Whether these assumptions are true or not is of small consequence in current research. The use of nonsense syllables persists because they are easily manipulated and controlled and actually form the basic material for investigating the effect of meaningfulness on learning. The learning of nonsense syllables probably requires no essentially different skill than does the learning of mathematical formulae, foreign language vocabularies, or the abbreviations for names of government bureaus. Facts based on nonsense-syllable learning can be applied to other situations.

2. **Meaningful material.** A great variety of meaningful material has been used, including lists of adjectives, nouns, and three-letter words of all parts of speech. Prose passages, poetry, and oriental language characters have been employed to some extent, although adequate control and analysis of such material is not easily accomplished.

Serial and paired-associate learning. We became somewhat familiar with these two distinct methods of presenting material in Chapter IX. As a brief review, let us put samples of the two methods in schematic form.

Paired Associates		*Serial*
vacant	..........→narrow	vacant
		↓
benign	..........→senior	narrow
		↓
discreet	..........→brittle	benign
		↓
pensive	..........→hardy	senior
	etc.	etc.

Note that in paired-associate learning the associations are "horizontal," whereas in serial learning they are "vertical." If paired associates are presented in different orders on successive trials, as is the common practice, no associations between *pairs* (vertical associations) are likely to be formed. On the other hand, an item in a serial list is first a response to a stimulus and then a stimulus to another response. *Narrow* is the response to the stimulus *vacant*, and *narrow* is the stimulus to the response *benign*.*

Methods of presenting verbal material. There are several ways by which verbal material may be presented to *S*.

1. *Memory drum.* This is the standard procedure used in most experiments (see Chapter IX).

2. *Card system.* For less exact work, *E* may type words on cards and present them successively to *S* at regular intervals, the intervals being indicated by a metronome or some other timing device. This method is well adapted for use in group laboratory work.

3. *Projection.* Some *E*s have used projectors to present material to a group of *S*s. The material may be printed on slides or strip films. This method allows for rapid collection of data from many *S*s, but restricts the nature of the material which can be used and the responses which can be measured. The use of the memory drum with individual *S*s is a laborious procedure, but for careful study of the learning process, and for securing complete records of errors and other observations, it has advantages which more than compensate for the labor. Over-all trends can be obtained nicely from group experiments, but for specific analyses the

* In serial learning the list is often preceded by a neutral symbol which is used as a cue for the first word in the list. This, in a sense, makes the first item functional since without the neutral symbol this first item would be used only as a stimulus for the second item.

memory drum technique, or at least the individual-S technique, is mandatory.

4. *Complete presentation.* Although used infrequently in recent years, this method is required for certain materials. The method of complete presentation using a list of nonsense syllables consists in giving S the entire list at once and allowing him a limited amount of time to study it. E then requests S to recall as many as possible and afterwards allows him another study period. This method is required for prose and poetry and other connected material to which the memory-drum technique cannot be adapted.

Maze Learning: Verbal-Motor

We now move up the dimension of motor activity. Maze learning, while actually placing less emphasis on the motor aspects of performance than is customarily believed, may be thought of as involving greater motor activity than does the usual verbal learning task. Mazes are of various kinds.

1. **Stylus maze.** In the stylus maze a path is cut into or indented in a base and the blindfolded S traces the pattern with a pencil or special pencil-like rod called a *stylus.* Two or more alternative paths are present at each choice point, only one of which is correct, i.e., will lead to the next choice point. The pattern may be cut through the base so that the actual tracings can be recorded on paper inserted under the maze. Metal plates (blind alleys) in circuit with counters may also be used to record the time spent in blind alleys or number of errors made.

2. **Raised or finger maze.** In this maze the index finger is used as the tracing instrument for a pattern which is raised or placed on top of a board base. The path is usually constructed of wire or small metal tubing.

The important factor in constructing the stylus and finger mazes is that the path be homogeneous throughout as far as feel or touch is concerned. This avoids the possibility that localized cues may influence the learning. The blind alleys in the finger or stylus maze may be of any number or length, but certain rather standardized forms have evolved, the most prominent being the multiple-U and the multiple-T patterns. In these mazes S has only two alternatives at each choice point, one alternative being right, the other wrong. Such mazes are shown in Fig. 63.

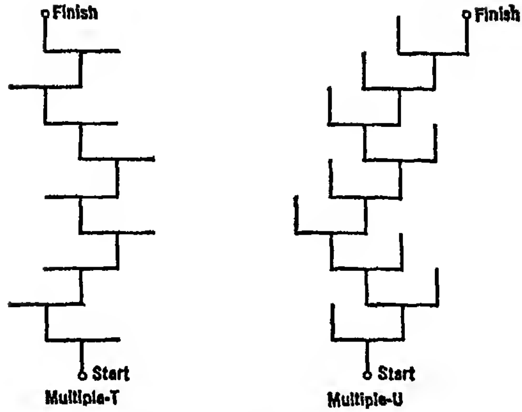


FIG. 63. Two common maze types.

3. **Multiple-choice type mazes.** There are several forms of this maze, but the most common is called the punchboard maze. Many rows of holes are bored in a large plaque which is placed before *S*. He is told that one hole in each row is correct and that he is to discover it by inserting a stylus into the various holes until some signal is given indicating correctness. When he inserts the stylus in incorrect holes, a red light automatically turns on, and when he puts it into the correct hole, a green light comes on. *S* continues through the series of rows, discovering in each case which hole is correct. Successive trials are given until *S* is able to go through the series without error. We may schematize a sample punchboard maze by letting O represent the five holes in a row, and x the correct hole:

Row 1	O	O	O	⊗	O
Row 2	⊗	O	O	O	O
Row 3	O	⊗	O	O	O
Row 4	⊗	O	O	O	O
Row 5	O	O	O	O	⊗
		etc.			

4. **Animal mazes.** In studies of rat learning, the multiple-U, multiple-T, or multiple-Y mazes have been used extensively, but since many mazes have been constructed to test specific hypotheses concerning the learning process, there have been scores of

different patterns. One need only leaf through a book largely devoted to animal learning in order to recognize that thousands of board feet of lumber have been nailed together in many different forms to test this or that hypothesis concerning learning.

Rat mazes may be of the enclosed-alley type in which the bottom and two sides are formed of boards with wire mesh for a top. Or, they may be elevated, in which case a single narrow board forms the path and is raised far enough off the floor to discourage the rat from jumping down. A third type is the water-maze, with which we are already familiar.

Maze learning has certain similarities to serial verbal learning, since a series of stimulus-response sequences must be integrated into a large pattern. *Unlike serial verbal learning, however, maze learning involves discovery of the correct responses.* In serial verbal learning *S* has only to fixate the correct response to a given stimulus, he does not need to discover what that correct response is. However, a few verbal discrimination experiments have been performed in which *E* simultaneously presents *S* with two or more words and tells him that only one of the words is correct. Within each of the several groups of such words *S* must discover the correct response. *E* must of course tell *S* when a choice is right or wrong; thus, the procedure becomes very much like punchboard maze learning.

Many tasks in addition to mazes have been used for the study of learning, tasks in which the skills may be said to be neither highly motor nor highly non-motor. None of these is used with sufficient frequency to be considered a standardized laboratory task.

Motor Tasks

Probably the only standardized device for studying motor learning is the *pursuit rotor*. In consists of a circular bakelite turntable very much like that of a phonograph. A small metal target disc about the size of a dime is imbedded an inch or two from the outside edge, so that its surface is flush with the surface of the turntable. The disc will, of course, move in a circular path when the turntable moves. *S* is asked to keep a metal stylus on the target, a task which is more difficult than it may appear since the stylus is jointed so that it must be held lightly on the target. When the stylus touches the disc an electrical circuit is

completed, thus activating a timer. The score is usually the number of seconds on target for a given interval of time (such as 1 minute), each interval being counted as a trial. *S* may alternately practice 1 minute and rest 1 minute. In some laboratories dual units have been constructed so that 2 *Ss* may be run simultaneously, with all starting and stopping of the rotors, as well as the scoring, being done automatically.

It will be noted that pursuit rotor learning does not actually constitute the learning of a series of responses to a series of discrete stimuli, as has been the case in tasks discussed just previously. There are part components in the task, although they are difficult to analyze into separate elements. Most motor tasks involve a series of coördinated movements which similarly are not easily analyzed.

METHODS OF RESPONSE MEASUREMENT

Performance is measured in many ways in multiple response learning, but these revolve around three primary methods with which we are already familiar: (1) mastery level after a given number of trials or for a given length of time; (2) number of trials or amount of time required to reach a given criterion of mastery set by *E*; (3) error reduction.

Learning Curves

In the previous chapter there were several illustrations of learning curves derived from conditioning procedures. In the typical learning experiment, response measures are recorded constantly so that any change in response acquisition may be seen from curves based on trial by trial records. Again we should mention that the shape of learning curves is a function of the conditions of the experiment as well as the transfer of learning from similar situations. So far as is known, there is no one learning curve—no "pure" curve, although in actual practice the negatively accelerated learning curve is most often found.

Learning curves for constant number of trials. When plotting a learning curve from data obtained by giving all *Ss* the same number of trials, or the same intervals of practice, there is no problem. Since we have a constant number of trials or constant

time, we merely average all scores for each successive trial or for each unit of time. These means give the best estimate of performance of the whole group at successive stages in learning. Since we shall have illustrations of learning curves for laboratory tasks later in the chapter, let us observe a curve for a non-laboratory task. Ghiselli and Brown (105) tallied the number of accidents of 60 street car motormen for the first 17 months that the motormen worked. The training time these men obtained before

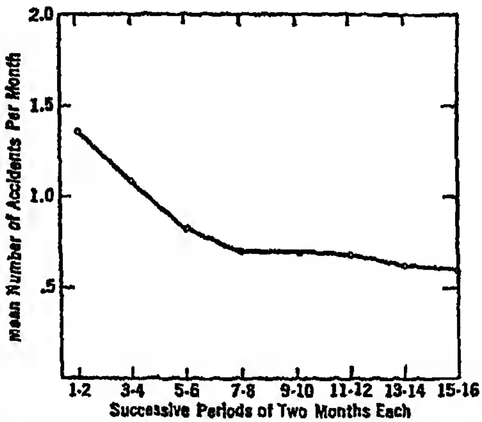


Fig. 64. Learning as measured by reduction in number of accidents attributed to 60 street car motormen. Estimated from Ghiselli and Brown (105).

starting on regular runs was never over one month and in most cases, only two or three weeks. The 60 men started work at various times during a 6-month period so that accidents which were a function of seasonal variations in traffic should have been fairly equally distributed over the 17-month period. Only a small portion of the reduction in accident rate is presumed to be due to improvement in motor skill as such. More important is the development of abilities to judge speed and distance, and to perform under stress.

The mean number of monthly accidents for the 60 "*Ss*" for the first 16 months of their regular runs is shown in Fig. 64. We have combined the frequency for successive 2-month periods. The curve has the customary negative slope.

In some instances the learning process may not be truly reflected

in the curve obtained by averaging scores for each trial. For example, let us suppose *S* is to learn a list of ten serial adjectives and we allot a constant number of trials, say 20, for learning it. Now suppose it turned out that by the end of 10 trials all *S*s have learned the list so they can say it correctly and for the next 10 trials they only repeat the list time after time. When we average *S*s' scores for trials 11 to 20, we would get a straight horizontal line, plotting successive trials against mean correct responses. In short, our graph might indicate that no learning had taken place on the last 10 trials, because there was a "ceiling" on the method of measuring the learning. Such a curve would not reflect the true state of affairs for those last 10 trials. Actually, the response strengths were almost certainly enhanced considerably by these 10 trials, a fact we could determine by giving retention tests. Therefore, in the event that we have a time or trial criterion, and any one *S* reaches apparent perfection of performance before the end of training, the curve of average scores will not adequately reflect the degree of learning.

Learning curves for mastery criterion. When performance has been measured in terms of number of trials or time to reach a criterion, some difficulty is encountered in plotting a learning curve. If we average the scores for each trial, we shall find that the fast learners soon drop out of the averages. As each *S* is eliminated (because we have no scores for him beyond his criterion trial), the N becomes smaller and smaller until for the last measure only one *S* may be involved. Obviously, by such a method, we should get a distorted picture of the relationship between trials and degree of learning. Some graphic way is needed to make the learning period equal for all *S*s; the fast learners' records must be stretched over a longer period of time and the period of the slow learners contracted.

Two methods for doing this have been worked out. One of them, the Vincent-Kjerstad method (150, 252), divides the total learning period into tenths, regardless of the length of the total period. By this method, each *S* is represented in each tenth, and the means for successive tenths give the picture of the learning function. As this technique involves considerable labor, another method, the trials-to-criteria method, is preferred. The computational labor involved in the trials-to-criteria method is somewhat

less than in the Vincent-Kjerstad method, and for most purposes the resultant curve is quite satisfactory.

In using the trials-to-criteria technique, one transformation from the raw scores must be made before the means for a group of *S*s are determined. Arbitrarily we set up successive performance levels and determine the number of trials (or time) it has taken each *S* to reach each performance level. Since all *S*s learn to a given final criterion, all *S*s will be represented at each performance level on the way to that criterion.

Let us illustrate the method by using data gathered by the author. *S*s learned a paired-associate list. There were ten pairs of adjectives in a list and each *S* learned to a criterion of one perfect trial, i.e., until he responded correctly to all the first members of the pairs on a single trial. The data for 10 *S*s are shown in terms of the number of correct responses given on each trial until the criterion was reached:

Trials *S*s	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
A	0	1	2	4	7	5	8	10								
B	0	0	3	3	3	4	5	7	8	6	7	9	7	10		
C	0	1	4	8	7	6	8	7	8	10						
D	0	5	7	9	9	9	10									
E	0	2	1	2	3	4	4	5	4	6	6	6	9	8	7	10
F	0	6	8	9	10											
G	0	2	7	6	6	8	9	10								
H	0	4	4	7	8	8	10									
I	0	4	5	6	10											
J	0	8	7	10												

Since the first trial is always a study trial, all *S*s have zero. One notes that there is considerable variation from *S* to *S* in terms of number of trials to learn the list to one perfect recitation. This is true even though these *S*s were well-practiced and were from a fairly homogeneous population. The individual scores show considerable irregularity from trial to trial and *plateaus* (periods of no apparent progress) are evident. For example, subject C scored eight correct on trial 4, but not until trial 10 did he show improvement beyond this.

From these raw data we make the transformation. Arbitrarily, we set up successive criteria which are the same for all *S*s, the criteria being one correct response, two correct responses, three

correct responses, and so on. To make a transformation we ask: "How many trials did it take *S* to reach one correct response? How many trials did it take him to reach two correct responses?" and so forth. Our successive criteria thus become one through ten correct responses, and we determine for each *S* the number of trials required to reach each criterion. For the 10 *Ss* this would be as follows:

Successive Criteria:	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Ss										
A	2	3	4	4	5	5	5	7	8	8
B	3	3	3	6	7	8	8	9	12	14
C	2	3	3	3	4	4	4	4	10	10
D	2	2	2	2	2	3	3	4	4	7
E	2	2	5	6	8	10	13	13	13	16
F	2	2	2	2	2	2	3	3	4	5
G	2	2	3	3	3	3	3	6	7	8
H	2	2	2	2	4	4	4	5	7	7
I	2	2	2	2	3	4	5	5	5	5
J	2	2	2	2	2	2	2	2	4	4
Means	2.1	2.3	2.8	3.2	4.0	4.5	5.0	5.8	7.4	8.4

Let us follow the first *S* all the way through this transformation. How many trials did it take *S* to get one correct response? The raw data show that he got one correct on the second trial so a two is entered. How many trials were required to get two correct? The data show he got two correct on the third trial. How many to get three correct? On the fourth trial he got four correct, so he also got three correct; consequently, we say that he got both three correct and four correct on trial four and an entry of four is made for successive criteria three and four. On the fifth trial we see that he got seven correct. This means, in terms of successive criteria, that he got 5, 6, and 7 correct on the fifth trial; a five is entered for these three criteria. To get eight correct it took *S* 7 trials; to get nine and ten correct it took 8 trials. In a similar fashion, other scores were derived for all other *Ss* so that each *S* is resprcsented at each stage of learning, where successive stages are defined as successive criteria, The means of the columns give the points for plotting the curve.

This curve has been plotted in two ways in Fig. 65. In the upper curve the successive criteria (independent variable) are placed along the abscissa and the mean trials to reach those criteria

(dependent variable) along the ordinate. This is the customary way to plot, i.e., with the variable measure along the ordinate. Note, however, that plotting in this fashion yields a positively accelerated curve which we know is not characteristic of the

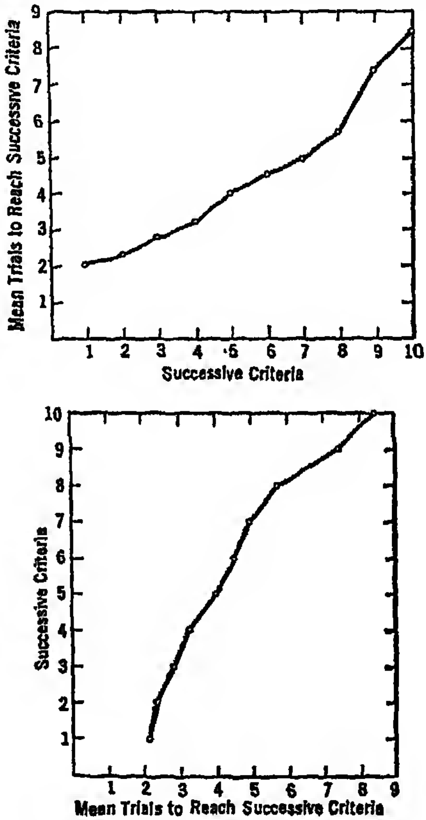


FIG. 65. Two methods of plotting trials-to-criteria learning data.

learning process. The curve as plotted indicates that on the average it took more and more trials to reach each successive criterion, which actually implies negative acceleration. The familiar picture of negative acceleration, however, is not given unless the axes are reversed as in the lower graph. Here, the trials to reach the criteria are placed along the abscissa and the successive criteria along the ordinate, thus yielding a negatively accelerated curve.

Both curves, of course, are accurate, and in actual practice both methods of plotting have been used.

The curves of Fig. 65 are somewhat irregular because of the small number of *Ss* involved. Also, they show that on the average it took more trials to go from a criterion of eight correct responses to a criterion of nine, than it did from nine to ten. This is atypical and would reverse itself with a larger number of cases.

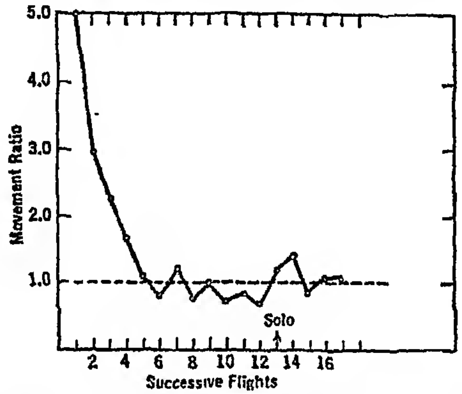


Fig. 66. The learning of elevator control in flying an airplane. The dotted line indicates proper control Estimated from Kellogg (191).

Individual learning curves. Many tasks, especially those of a motor nature may require many practice periods before a high level of performance is reached. If these tasks are such that discrete parts of the performance may be measured separately, it is often worthwhile to plot learning curves for each part for a single *S*. By doing this, a study can be made of the sources of difficulty in learning the task. During World War II great interest was focused on skills required to fly an airplane, and attempts were made to measure objectively the movements of the ailerons, elavator, and rudder independently. If the correct movements of these controls could be compared with the movements actually made by a student, the direction and extent of error could be plotted from flight to flight.

A device for measuring the actual extent of control movement, developed at Indiana University, was used by Kellogg (191) to study the progress of learning of student pilots. The learning

curve for elevator control for one *S* is shown in Fig. 66. Each point represents 10 minutes of recording of elevator movements while *S* was flying over a standard course which included four left turns and three right turns. The correct movements of the elevator were determined by recording these movements as a skilled pilot flew over the same course and under the same weather conditions as those under which the student had flown. Dividing the mean extent of elevator movement of the student by that of the skilled pilot, gave an index of error. Thus, a ratio of 1.0 would indicate movements which were exactly the same, while ratios greater than this would indicate that *S* erred by moving the elevator's controls excessively.

Figure 66 is the learning curve for elevator control for approximately 17 flights of 30 minutes each, plus the 10 minutes needed to record the movements. A curve of negative slope is apparent. As far as elevator control was concerned, this student was performing quite well by the end of the fifth lesson. Only a slight increase in error was apparent on the first two solo flights.

VARIABLES INFLUENCING RATE OF LEARNING

Massed vs. Distributed Training

We are already acquainted with the general methods of manipulating this variable. In the previous chapter wherein these methods were discussed, data were cited showing conditioning taking place more rapidly under distributed than under massed practice. In general, the same conclusion holds when tasks are of the multiple-response type. McGeoch, after an exhaustive survey of the literature, says: "The generalization that some form of positive distribution yields faster learning than does massed practice holds over so wide a range of conditions that it stands as one of our most general conclusions" (233, p. 119). We shall see, however, that there are exceptions to the generalization.

Let us examine some experiments dealing with this variable as it is related to specific sub-variables, and while we do this, sample methods at various points along the verbal-motor continuum.

Verbal tasks. In a series of experiments Hovland has attacked a group of problems associated with massed and distributed practice,

consistently using nonsense syllables as the learning material. To illustrate: one experiment studied rate of learning under massed and distributed practice when the syllables were presented at different rates (158),—a 2-second exposure rate (typical rate) and a 4-second rate of presentation.

We can orient ourselves to the *S*'s point of view by quoting from the instructions which Hovland gave to his *S*s, since these are representative of instructions used in rote learning experiments:

> This is an experiment in learning a list of nonsense syllables, and not a psychological test. We are interested in certain complex relationships of the learning process common to all people, and not concerned with your personal reactions.
>
> Shortly after the apparatus starts you will see a three-letter syllable in the window. You are to pronounce this syllable and those that follow it as you see them. After you have seen the list once, you are to endeavor to anticipate the syllables; in other words, as you see one syllable you are to pronounce the syllable that will follow it *before* it appears. If you think you know what a syllable will be, but are not sure, guess, because it will not hurt your score any more than to say nothing, and if you get it right, it will count as a success. If you anticipate a syllable incorrectly, correct yourself as soon as it appears. Try always to speak the syllables as distinctly as possible.
>
> Please do not try to think ahead more than one step at a time, or to count, or to make up fanciful connections between the syllables to assist the learning process. Don't try to use any special system in your learning; simply associate each syllable with the next one as the series moves along (157, p. 205).

Massed training was defined by a 6-second rest between trials and distributed practice by the 2-minute 6-second interval. With the addition of the presentation-rate variable, we have a total of four conditions:

Condition I: Massed; two-second presentation rate
Condition II: Distributed; two-second presentation rate
Condition III: Massed; four-second presentation rate
Condition IV: Distributed; four-second presentation rate

Each *S* in this experiment went through each of the four conditions twice, so that with 32 *S*s, each condition is represented 64 times. Systematic randomization (Design Method IV) was used to equalize the practice effects for the eight experimental sessions,

and two practice lists were learned before the experimental conditions started. Each list was composed of 12 syllables.

The learning curves for each condition are plotted in Fig. 67 We note first that with a 4-second presentation rate the lists were learned much more rapidly than with a 2-second presentation rate. Roughly, learning was twice as fast with a 4-second rate as with a 2-second rate, suggesting that amount of learning per unit

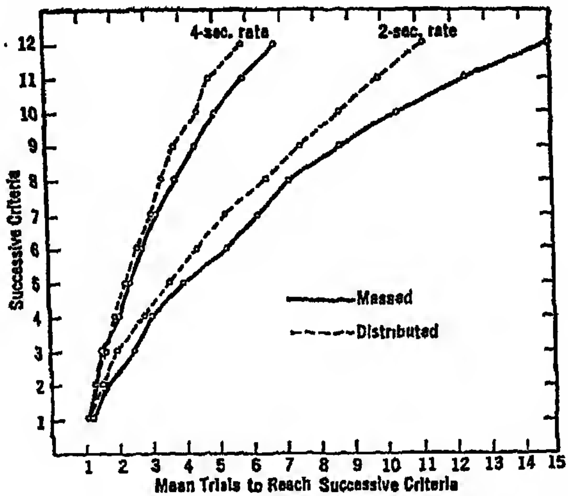


FIG. 67. The influence of massing and distribution under two rates of syllable presentation. Data from Hovland (158).

of time was about the same for the two conditions. With regard to massed vs. distributed practice, we see that distribution facilitates learning in both instances, although the difference is significant only in the case of the 2-second rate. We must conclude that the intra-trial distribution provided by the 4-second rate was not further enhanced an appreciable amount by also allowing 2 minutes, 6 seconds between trials.

Using the same basic methodology outlined above, Hovland has found that the longer the list of syllables the greater is the efficiency of learning resulting from distribution of practice (161), and that distributed practice with paired associates does not facilitate learning (159). This latter finding again emphasizes differences in learning paired associates as compared with serial

lists, and shows that distribution does not always enhance learning.

Wilson (415) has found that the learning of 16-item serial lists of two-syllable adjectives will be facilitated by using either a 30-second or 1-minute rest between trials. However, in a study by Sandahl (329), in which serial lists of adjectives were used, no such facilitation was found in several different conditions. Let us look at Sandahl's conditions:

Condition I:	1-min. rest after each block of 2 trials
II:	2-min. rest after each block of 2 trials
III:	4-min. rest after each block of 2 trials
IV:	1-min. rest after each block of 4 trials
V:	2-min. rest after each block of 4 trials
VI:	4-min. rest after each block of 4 trials
VII:	Massed

Following two practice days these seven conditions were presented to each of 21 *Ss* by Design Method IV. Each list consisted of six pairs of highly synonymous adjectives, for a total of 12 items in a serial list. In no case did a word immediately follow or precede its synonym. The purpose in using lists with high intra-list similarity was to increase the number of errors made during learning, since evidence has tended to suggest that distributed practice will most facilitate learning when interference within a list is high (167). Sandahl's results show that whereas there was considerable interference within the list, there was little difference in the speed of learning among the results of any of the above seven conditions. This is one of the few known instances in which distribution has failed to facilitate learning of serial verbal material. It may be that these conditions are not optimum for distribution effects; perhaps 1 minute or less rest between each trial, as used by Wilson, is the maximum interval for which distribution proves beneficial in serial learning. Or, it may be that the use of synonymous adjectives, while producing interference, added factors which obscured the usual benefits of distribution.

Semi-verbal materials. As we go up the verbal-motor dimension toward the motor end, we find that distribution of practice facilitates performance even more than it typically does in the learning of verbal materials. We shall mention two illustrations of tasks which fall somewhere on the dimension between "pure verbal" and "pure motor."

Ericksen (90) used a multiple-U stylus maze with ten choice points. *S*s in one group learned the maze during a single period, and *S*s in another group were given 2 trials a day until a criterion of two successive perfect runs was reached. The group working under distributed condition required an average of only 14.06 trials to reach the criterion and made a mean of 35.65 errors. The massed group took an average of 28.21 trials and made an average of 70.0 errors. Thus, the distributed group took approximately half as many trials to learn as did the massed group. The two groups of *S*s were not matched, but even allowing for differences in initial ability of the two groups, the obtained differences are so large that it seems clear distribution had a very facilitating influence. Let us note that 24 hours elapse between blocks of 2 trials in the distributed condition. We do not know, of course, that this particular work-rest ratio is optimum. Other conditions of distribution might accomplish even more rapid learning. We can see that a great many combinations of work and rest must be explored before *E* can definitely state the optimal conditions of distribution for any given task.

A second learning task which fits the present category is learning to print the alphabet upside down and backwards. *S* starts at the right side of the page and prints A upside down, then B upside down, and so forth, so that when the page is turned 180 degrees, the alphabet can be read in sequence. *S* is instructed to print as rapidly as possible. This task is good as a class demonstration experiment, since it can be done with groups of *S*s and performance almost always shows marked facilitation as a result of distribution. We shall illustrate this by Kientzle's study (196).

*S*s were given work periods of 1 minute. If they finished the alphabet before the period ended, they immediately started over. A different group of *S*s was used for each time interval, this particular task being of such a nature that it cannot be used with counterbalanced conditions effectively (see Chapter X). The groups ranged from 35 to 63 *S*s and the data show that in terms of initial scores all groups were of near-equal ability. Ten degrees of distribution were used, as shown in Table 25. It will be seen that the intervals for the nine groups vary from 0 to 90 seconds and that an interval of 7 days between trials was given to the tenth group.

In general the results show that as the time interval between rials increased up to 90 seconds, the rate of performance increased lso. Even the group which had one week between trials did much etter than did the short-interval groups.

TABLE 25

EFFECT OF DISTRIBUTION OF PRACTICE IN REVERSED-ALPHABET PRINTING

	Interval Between Trials (Sec.)									
	0	3	5	10	15	30	45	60	90	7 Days
Mean Scores on Tenth Trial	42.2	48.3	48.4	49.2	53.6	56.1	59.2	55.9	59.9	56.8

Data from Kientzle (196).

The Kientzle experiment also permitted isolation of the effects of rest-interval activity. These effects will be discussed after a few general observations on this variable are made. In studying the effects of short rest intervals on learning, *E* must keep *S* in the laboratory. This raises the problem of what to do with *S* during the rest interval. Any rest interval activity must meet two criteria if we are to get out of its use a good estimate of the influence of distribution as such.

(1) *A rest interval activity must prevent S from rehearsing the learning material.* If we get facilitation by distribution, and if *S* has spent the interval silently practicing, then we are not truly measuring the influence of the rest interval. Consequently, it has been common practice to give *S* a task which keeps him occupied. However, certain pitfalls are involved in this.

(2) *A rest interval activity must not interfere with or facilitate the learning of the material.* If the rest interval activity interferes with the learning material, the influence of distribution will be underestimated; if the rest interval activity facilitates, the influence of distribution will be overestimated.

Es have used many rest interval activities. One commonly employed in verbal learning is color-naming. *S* spends the rest interval naming the color of many small paper squares pasted on a board or squares that are presented by the memory drum at regular intervals. This, it is believed, does not interfere with the learning

material and at the same time seems to keep *S* occupied so that he is unable to rehearse. Such an activity is especially suitable for short rest intervals. For longer intervals *Es* have used jigsaw puzzles, picture magazines, jokes, cartoons, and so on. So far, little empirical attention has been paid to the choice of materials; instead, *E* uses his judgment as to whether or not a rest interval task will interfere or facilitate and whether or not it will keep *S* busy so that he cannot rehearse. It has been shown that if *S* is requested to practice silently during the interval, his performance

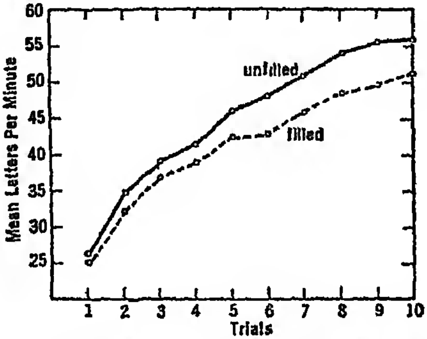


FIG. 68. The effect of filled and unfilled rest intervals on reversed alphabet printing. Data from Kientzle (196).

will be facilitated on subsequent learning (416), so it is necessary that *S* be prevented from rehearsing.

Now, to turn back to Kientzle's study. During the rest intervals between trials in learning to print the alphabet backwards, "*Ss* were not to look at their papers and not to think or talk about the task. So far as *E* could detect, these instructions were followed" (196, p. 190). In a special group of *Ss* (not included in the table), *E* gave 60 seconds between trials and had *Ss* make tallies and crossbars throughout this interval. For this group, then, the interval was occupied. Let us compare the performance of this group with the performance of the 60-second group in which the interval was not formally occupied (Fig. 68). We see that the group with the interval unfilled did consistently better than the group which filled the interval by making tallies and crossbars, although the groups are large (54 and 55 for filled and unfilled groups, respectively), and are about equal in performance on the first

trial. By the tenth trial the difference between the two groups is such that it could be expected by chance only once in fifty times if the true difference were zero. How do we account for this difference?

Several alternative hypotheses present themselves as explanations of the difference between the two groups. (1) The task selected to fill the interval interfered with the printing of the inverted alphabet. Such an interference might be in terms of actual movements or might be in terms of accumulation of fatigue. (2) *Ss* may have been practicing the alphabet in some fashion when the interval was not filled. This would tend to enhance performance and there is independent evidence that *Ss* may symbolically rehearse even a highly motor task (307). (3) There is no true difference between the groups. Even though the curves are consistent throughout their course, it is still conceivable that slight differences in the groups at the start represent true differences, although this is not a probable state of affairs. From the data as given there is no way of selecting which is the correct hypothesis, although the first, the effect of interference, seems most likely. The point to stress is that *how we fill the rest interval is important in distributed practice studies if we are to get an unbiased estimate of the influence of distribution of trials as such.* It should be clear that the results for these two groups do not negate Kientzle's general findings, since the differences between the other groups are too great to be accounted for by the rehearsal hypothesis as proposed above.

Motor task. No available evidence contradicts the general principle that some form of distribution facilitates performance on a motor task. One study of the pursuit rotor will show method and results.

Hilgard and Smith (154) kept the practice period constant for three groups of *Ss*, but varied the number of trials given during the period. All practice trials were 1 minute in length, and all groups were given two initial trials with 1-minute rest intervening. From the third trial onward the groups were distinguished by the amount of rest between trials as follows:

Group A: 5 additional trials with 3-minute rest between trials
Group B: 10 additional trials with 1-minute rest between trials
Group C: 15 additional trials with 20 seconds rest between trials

All groups were given a final trial 1 minute after their last trial above. Counting the first two trials given all groups, the total experimental period for each group was twenty-five minutes. Group *A* was given relatively distributed practice as compared with Group *C*, and Group *B* was intermediate.

The pursuit rotor was wired so that during a 60-second trial a maximum score of 600 could be made; that is, if *S* succeeded in

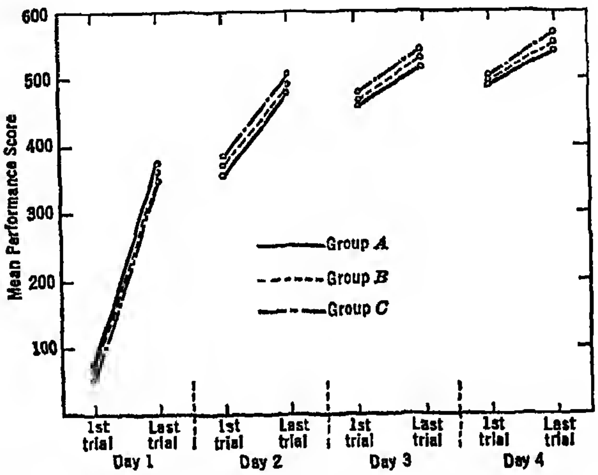


FIG. 69. Massing and distribution of trials in pursuit-rotor learning. Data from Hilgard and Smith (154).

constantly holding the stylus on the target, he would score 600. Each group was given four experimental periods exactly as outlined above, and the scores made on the first and last trials for the periods are shown in Fig. 69. It must be kept in mind that the last trial of a day for Group *A* was actually the 8th trial; for Group *B* it was the 13th trial, and for Group *C*, the 18th trial.

Figure 69 shows that the three groups performed almost equally well throughout the four days. By the end of the fourth day Group *A*, which had had a total of only 32 trials, was doing almost as well as Group *C* which had had 72 trials. In short, the fewer trials within the constant length practice period resulted in nearly as much learning as the greater number of trials for the same period.

Further problems are associated with massed and distributed practice, problems which concern the optimum point in learning for introducing a single rest interval and the optimum length of rest interval at various stages of learning. For our purposes, let us keep in mind the general finding that distributed practice is superior to massed practice, and that how much better distributed training is than massed depends upon three general manipulable variables:

1. Relative and absolute amounts of work and rest—the work-rest ratio
2. Locus of the rest period or periods
3. Type of material or task used

Theory. Most theories attempting to explain the facilitation of distributed training over massed training have as their bases the postulated opposition of two kinds of processes, one inhibitory and the other excitatory, with differences in the rate of dissipation of the processes related to passage of time. It is, then, based on the same premises as the theory of distributed practice discussed in the previous chapter. The essential idea is that rest intervals allow for the dissipation of inhibitory processes, thus leaving the positive tendencies stronger. Actually, the data do not appear to be explained by such a simply stated theory, and recent trends are toward more complicated theories, although even these are based on the essential idea of the opposition of the basic processes.

Type of Material

Three specific variables determine how the type of material affects rate of learning. They are: (1) intra-task similarity, (2) meaningfulness, and (3) affective characteristics.

Intra-task similarity. Although this variable has received little experimental attention, it is probably an extremely important one in determining rate of learning. It might be stated in terms of degree of homogeneity or heterogeneity of the items or units of a task, but to make the analysis consistent with our previous discussion of transfer of training, we should retain the concept of similarity as the stimulus variable in question. It is believed that the similarity factors among the items *within* a list, or movements in a task, operate in the same manner to produce negative transfer

as does similarity *between* lists. Within limits, *the greater the similarity among parts of a task, the greater is the negative transfer within the task, hence, the slower is the learning.*

Gibson (108) used paired-associate lists constructed with figural designs as stimuli and nonsense syllables as responses. These materials were the same as those used in her study reported in Chapter IX. Each list consisted of 12 pairs. One list was constructed with stimulus forms quite dissimilar, and three others with highly similar stimulus forms. Similarity was achieved by taking four of the stimulus items of the dissimilar list in combination with two highly similar forms for each. Thus, each of the three similar lists consisted of four sets of forms, each set highly similar within itself. In considering the results we shall evaluate performance on the dissimilar list as compared with mean performance on the three similar lists combined.

A different group of *S*s was used for each condition, and the criterion of learning was one perfect recitation. The dissimilar list required a mean of 8.86 trials to learn, whereas the similar lists required a mean of 19.81 trials—over *twice as many as did the dissimilar list.*

Gibson interprets her results in terms of intra-list generalization. The stimuli of the similar lists tend to generalize with each other and evoke many wrong responses—responses which actually should be associated with other stimuli (response generalization was probably at a minimum). In learning the dissimilar list *S*s made an average of about three errors during learning, whereas in learning a highly similar list an average of about 18 errors occurred. Clearly, there was a great deal of interaction of the stimulus members of the similar list.

The errors which occur in Gibson's experiment constitute the basic empirical evidence for generalizing tendencies. When an item evokes a response which is wrong, and when it can be seen that the wrong response is actually the correct one for a stimulus item which is highly similar to the one which elicited it, the evidence for generalization is apparent. If we plot errors throughout the course of learning, it can be shown that their frequency increases to a maximum at about one-third of the way to mastery and then gradually decreases. Gibson's error curve shows this trend as have other curves using different types of material (329).

We conclude that slower learning occurs with lists of similar items because generalization is high. In order for learning to occur, the generalizing tendencies must be reduced by differential reinforcement in the manner previously discussed. Lists of low similarity required much less of this differentiation.

The work on code learning during the war has given additional information concerning the importance of intra-task similarity. Studies by Keller and his students (189, 190) will illustrate these findings. There are 36 characters in the International Morse Code, 26 for the letters of the alphabet and 10 for numbers zero through nine. The 36 characters were learned by presenting the telegraphic signals for them over and over, with *S* attempting to identify the characters. Immediate correction of errors was made by *E* and the order of presentation varied in random fashion from trial to trial. The 36 characters were then divided into four groups depending on difficulty encountered during learning. It was shown that difficulty of the four groups was directly related to the number of errors made during learning. That is, characters which were difficult to learn tended to be substituted for each other during learning, whereas the group of characters which was learned rapidly produced few such errors. These findings suggest that intra-group generalization is producing the differences in difficulty.

Plotkin (292) went further to show specifically the relationship between similarity of symbol groups and difficulty in learning. This was accomplished by scaling similarity of characters first, and then making up lists of different degrees of similarity while holding constant individual item difficulty as such. His results clearly show that as similarity among items in a list increases, the time to master the symbols in a list increases; generalization increases the interference.

The concept of generalization has also been used by Phillip and Peixotto (291) to explain the errors they observed in paired-associate learning of nonsense syllables. Cofer (61), after analyzing the errors made in memorizing prose passages, suggests that interference brought about by certain similarity factors will probably account for most of the errors. Thus, *intra*-task negative transfer may be interpreted in much the same way as *inter*-task negative transfer. We could, of course, by the use of similarity

relationships within a task, make our conditions such that intra-task transfer would be positive in nature.

Meaningfulness. Like intra-task similarity, meaningfulness is a very important variable in determining the rate at which learning takes place although, again, the amount of experimental work on the variable is meager. Meaningfulness for verbal units may be defined operationally in two ways:

1. It may be the number of associations which a word elicits in a limited period of time. Thus *E* presents a word as a stimulus and asks *S* to write or name all other words or ideas of which this word makes him think. The greater the number of associations elicited the greater is the meaningfulness.

2. It may be the per cent of *S*s who give an association (word or idea) within a limited period of time. The greater the number of *S*s who can give an association the greater is the meaningfulness.

Published lists of nonsense syllables (112, 164, 206, 417) are reported in terms of association value which is determined by the per cent of *S*s giving an association within a limited period of time. Thus, syllables which have an association value of 100 per cent indicate that all *S*s reported an association or idea within the time limit allowed. Let us take a few illustrations from Glaze's list (112):

100 per cent	*53 per cent*	*0.0 per cent*
DOB	*BOZ*	*DAX*
PIL	*KER*	*ZYD*
SYN	*LYB*	*XUC*
POK	*WOB*	*JYC*
HOS	*QAT*	*WUH*

You will probably agree with the *S*s from whose responses these values were derived that the 100 per cent association value syllables quickly suggest other words or ideas, whereas the 0.0 per cent syllables are much less suggestive. Actually, because of cultural changes, meaningfulness values of syllables should probably be checked or re-computed from time to time. For example, in Glaze's list (compiled in 1928) the syllable *DUZ* appears with an association value of 47 per cent. Very likely the value of this syllable is 100 per cent among current college students.

The relationship between meaningfulness and rate of learning has been determined by the use of nonsense syllables of different

degrees of association value. In so far as the work which has been done is concerned, the data are unequivocal; *the greater the degree of meaningfulness the more rapid is the learning.* Briefly, let us look at the results obtained by McGeoch (231) on this problem.

By the method of complete presentation, 98 *Ss* were presented four lists of ten items each. One list was made up of three-letter words, another of 100 per cent association value syllables, another 53 per cent, and a fourth, of 0.0 per cent syllables. Each *S* learned all four lists, a form of systematic randomization being used to balance practice effects. Rather than learning to a criterion, each *S* was asked to recall as many of the words of a given list as possible after having studied the list for a given interval of time. The results are shown in Table 26 in terms of mean number correctly recalled after a constant study period for each kind of list. There is a direct relationship between meaningfulness and amount learned.

TABLE 26

LEARNING AS A FUNCTION OF MEANINGFULNESS

Nature of Material	*Mean Number Items Recalled*
0.0% syllables	5.09
53.0% syllables	6.41
100.0% syllables	7.35
3-letter words	9.11

Data from McGeoch (231).

Meaningfulness as a determinant of rate of learning probably gains its status as a consequence of positive transfer. It is the same characteristic as implied by the term *familiarity.* Meaningful material allows *S* to draw on a vast store of existing familiar associations, in order to facilitate learning. The mechanics by which this associative process takes place are quite obscure. In certain situations it appears that a process of mediate association takes place (Chapter XI). In others it seems to be a form of generalization. Thus, if *S* has to learn to say *MON* when he sees *BAL,* he may knowingly generalize to think of BALl—dia*MON*d. Such associations are often reported by *Ss* although it is very difficult to get systematic data on them. All that we know is that material which is barren of associative connections will be learned less rapidly

than material rich in associative value. We can, of course, devise situations in which previous associations interfere with new learning, but as we have seen in the chapter on transfer of training, these situations are restricted in scope.

The practical value of meaningfulness in the teaching situation is widely recognized. Frequently, in introducing new concepts, a teacher will attempt to orient the student to the new concept by indicating its relationship to known concepts. Descriptively, such a student is provided something to which he may tie the new concept.

Affectivity. In the chapters on the discriminal processes we studied illustrative experiments in which *S*s scaled words for affective value. It was noted that many words had rather universally unpleasant connotations, and many other words had rather universally pleasant connotations. Unpleasant words tend to evoke avoidance responses; pleasant words, approach responses. Words with neutral affective tone evoked neither approach nor avoidance responses. The question we ask now is whether or not affectivity is related to the speed of learning.

Unfortunately, this question cannot be answered. Various studies have produced contradictory results apparently for the reason that dimensions other than affectivity have been inadequately controlled. Let us outline the steps necessary to determine adequately the influence of affectivity on rate of learning.

1. Scaling of words for affectivity. *S*s must be clearly informed as to what is meant by affectivity, and the scaling procedure should include scaling of the same words along other dimensions to determine if affectivity is an independent dimension. For example, meaningfulness would have to be scaled. Other dimensions by which words might possibly vary should also be investigated independently. The dimension of impressiveness as used by Thorndike (378) may be the same as affectivity, but at least an empirical investigation of whether it is or is not a separate dimension should be undertaken. In short, *E* should ask *S*s to scale a given group of words for all dimensions by which there is reason to believe these words may differ. This is an absolutely necessary step. From this scaling we presume that we could derive dimensions of affectivity, meaningfulness, and perhaps others. It is barely possible that meaningfulness itself may be systematically

related to affectivity (or vice-versa) in which case we would not have evidence for an independent dimension.

2. The second step involves the construction of word lists of varying degrees of affectivity; *all other dimensions by which the words vary are held constant.* Thus, meaningfulness would be constant for all lists while affectivity varied. The affectivity ratings of the words used must have high reliability—must be unambiguous in affective value. At the same time we would hold constant for all lists the factors introduced by making lists from individually scaled items. The important variable to control in this instance is intra-list similarity, which we have seen profoundly influences rate of learning. Consequently, when a list is constructed, the degree of intra-list similarity must be constant for all lists. The ways in which similarity may vary were referred to in the chapter on transfer of training. Synonymity, homonymity, antonymity, formal similarity in terms of prefixes, suffixes, and so forth, must be constant for all lists. We repeat: *only affectivity must vary.*

3. With these laborious but necessary preliminary steps completed, the word lists may be presented to *S*s in a learning experiment. The sample of *S*s used in the experiment should be drawn from the same population as was the sample used to scale the affective tone of the words. Affectivity would certainly be expected to be determined culturally, and different groups may have different affective responses to the same stimuli.

Until such procedural steps are followed, the results regarding affectivity and rate of learning will remain ambiguous.

In the discussion of the three variables in this section our illustrations have been necessarily studies of verbal learning, since we have no experiments on motor learning in which these variables have been manipulated. We can conceive of motor tasks in which similarity among movements may be varied, and it is likely that the intra-task interference effects would be much the same as those with verbal material. It does not seem possible at present to define and vary meaningfulness for motor tasks. The same is true concerning affectivity.

In addition to the three dimensions discussed, there may be others along which material may vary. The degree of motor involvement, for example, may be a variable; the amount of

discovery of the correct response required by the task may be another. However, the three considered are the only ones of importance which have had experimental attention. Meaningfulness and intra-task similarity are extremely critical variables; affectivity is a potential variable. An understanding of experimental problems involved in investigating affectivity is fundamental since these problems are common to all studies in which type of material is to be varied.

Knowledge of Performance

In the typical set-up for a verbal learning situation *S* can perceive immediately whether his responses are right or wrong. There are some tasks, however, in which such knowledge of the adequacy or inadequacy of responses is difficult for *S* to secure. For example, in Chapter II we had an illustration of direction perception in which *S* was to set a small dot on a screen so that it was in the same plane as a line which was also shown on the screen. Performance on this task was much better if *S* was told after each setting how much and in what direction he "was off" than if he were not told this.

We shall examine three experiments in which this variable has been studied.

Manipulating a gun sight. *Ss* were given a six-session training program on a laboratory model of a gun sight (344). The target was a slide-projected silhouette of a plane moving across a circular screen. Operating the gun sight involved two skills, tracking and ranging the target. Manipulation of hand dials regulated the diameter of a projected circle outlined by several diamond-shaped points (pips) of light. The target was correctly ranged when the circle outlined it, the wings forming the diameter of the circle. Tracking the target consisted of maintaining a point of light in the center of the circle on the nose of the plane. The skill involved in tracking was a relatively simple one and *S* could tell immediately whether he was tracking correctly, i.e., when the dot was on the nose of the plane. In the case of ranging, however, the adjustment required was precise, and it was difficult for *S* to tell when he deviated slightly from the position which was scored a success. Since *S* was required to track and range simultaneously, and since the target plane moved at considerable speed,

the task as a whole was rather difficult. Scoring was in terms of mean time on target for eight "runs" of the plane across the screen, the scoring being independent for tracking and ranging. Each session consisted of 64 runs.

One group of 20 *S*s practiced six days on this task without receiving any information as to the adequacy of their responses except in so far as they could inform themselves. This information was probably quite complete in the case of tracking, and quite

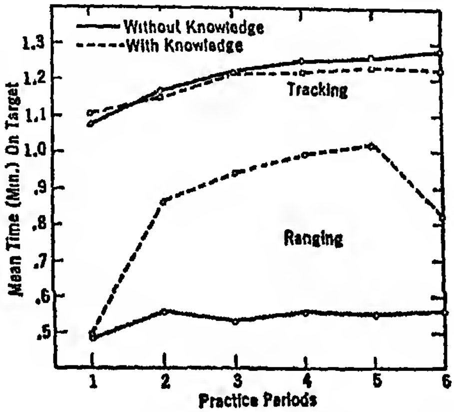


Fig. 70. The effect of knowledge of performance on the operation of a gun sight (344).

incomplete in the case of ranging. Another group of *S*s also practiced six days, but from the second through the fifth days a red filter dropped in front of the projector beam whenever *S* was both tracking and ranging correctly. Thus, the *S*s in this group were given a definite "statement" of when they were right and when they were wrong; when the plane image was red, they were "right," when it was not red, they were "wrong." Knowledge of performance was immediate and unambiguous.

The mean performance of each group for the six days is shown in Fig. 70. There is little difference in the tracking scores of the two groups. For ranging, however, the results are quite different in that knowledge of performance enhanced scores a great deal. We may note also that the scores for the group given information dropped sharply on the sixth day when the information was

withheld. Clearly, performance was directly related to knowledge of that performance which *S* obtained.

Degree of knowledge of performance. The experiment with the gun sight was such as to give complete knowledge or no knowledge. We would predict that the more precise the knowledge of performance the faster is the learning. Such an expectation is verified in an experiment by Trowbridge and Cason (389). *S*s, while blindfolded, were instructed to draw three-inch lines. One group of *S*s was told nothing after drawing each line; another group was told "wrong" if its errors were large, and "right" if its errors were small, and a third group was told exactly how much and in what direction it erred. Accuracy of drawing was directly related to amount of knowledge given.

Industrial training. New employees in most industrial jobs of course require some training before they are considered competent. Lindahl (221) has studied the effect of knowledge of performance in the industrial training programs for disc cutters.

Disc cutting in this particular plant involved a rather precise coördination of the hands and one foot. The discs, cut from tungsten rods, were very thin pieces to be used in electrical apparatus. The wheel which cut the discs from the rods was .015 inch thick, and was not visible to the operator during the actual cutting. The right foot controlled the cutting wheel; learning the correct foot pressure to be applied at various points in the cutting sequence was the most difficult part of the task. One hand was used to insert and hold the tungsten rod, and the other hand to manipulate a lever which ejected the discs after cutting.

Lindahl attached a device to the cutting machine to record the exact foot movements made by the operator. By taking several samples of movements from skilled operators, he was able to develop an ideal pattern of foot movement for controlling the cutting disc. Then by recording movements of new trainees, he could point out to these trainees just how their movements deviated from the correct movements. Instructional sheets showed the trainee just what damage to the discs was caused by the various kinds of incorrect movements and pressures. Special emphasis was called also to movements which would avoid breaking the expensive cutting wheels or wearing them down too rapidly. Records were taken on several days so that the trainee

was able to watch closely how his pattern of movement improved with respect to the correct movement pattern.

The effectiveness of the knowledge of performance given the trainees was determined by comparing the performance of the trainees with the performance of (1) veteran operators who had picked up the skill in a "hit and miss" fashion, and with the performance of (2) five new operators who were learning the skill by the "hit and miss" method. These comparisons were admittedly crude since the number of cases in the various groups was small and there was no way of knowing whether or not these small groups were of equal "natural" ability before training. The training program lasted 12 weeks. In terms of number of discs cut the effect of the training program was negligible. However, in terms of number of discs cut per wheel, and in terms of number of wheels broken, the training had a clear-cut effect. The trained group broke fewer wheels and obtained more cuts per wheel than did the formally untrained groups. Since these wheels were expensive, a knowledge of performance training method had high practical significance.

How do we account for the consistent finding that knowledge of performance enhances that performance? There are at least two ways by which this variable is believed to increase the rate of learning. First, *S* is given information concerning the incorrectness of his responses. Without this information, he may continue to repeat these incorrect responses. Knowing which responses are incorrect, however, *S* will change his behavior in an attempt to discover the correct ones. Thus, we may say that knowledge of incorrect responses increases the speed with which the correct response is discovered. Secondly, the knowledge of performance probably increases or at least maintains motivation.

Miscellaneous Variables

In this section we shall call attention briefly to a few additional variables.* The brevity of treatment is justified for three reasons: (1) few methodological problems are present in experimentation

* The most comprehensive surveys of all variables are given by McGeoch (233), Kingsley (198), and Stroud (370). There are many studies concerned with "best" method of presenting a specific kind of material. Since the conditions manipulated are often specific to the material, the results do not have general significance.

on these variables; (2) some of them are relatively ineffective variables in that they do not markedly influence the rate of learning, or (3) currently, very little research is being done on them.

Whole vs. part learning. If a task may be divided into discrete parts, it is pertinent to ask whether learning each part as a unit before putting the parts together will yield faster or slower learning than practicing on the whole from the outset. Thus, if we have a 10-stanza poem to memorize, will it be better to master each stanza separately before trying to recite the entire poem or better to attempt learning the poem as a whole from the beginning? There can be several variations on the pure whole method and the pure part method. For example, perhaps two stanzas of the poem might be considered a part; perhaps we might learn stanza one and stanza two and then try reciting them together before moving on to the third.

For ten years there has been little research on this variable. A stagnation of research on a problem usually indicates that (1) no further researches are needed—the answers to questions are known; (2) experimentation has reached an impasse in that results have been contradictory for no apparent reason, or (3) there is a lack of adequate theory to direct experimentation. We know that the first alternative is not true—there are many questions needing answers. The second and third alternatives combined seem to account for the state of affairs concerning this variable. Results have been contradictory, sometimes showing facility in learning by wholes and sometimes by part, and psychologists have been unable to specify adequately the nature of the differences between experiments which would account for discrepancies. The reason for this is probably that no comprehensive hypothesis concerning the variable has been set up by which the success of the studies may be evaluated. If there is a cluster of conditions which must be present before whole learning is superior or a cluster of conditions which must be present before part learning is superior, we have not isolated them. If the relative effectiveness of one method over another is shown only under highly restricted conditions, as is probably the case with whole and part learning, it is a relatively minor variable.

Active recitation. If during learning *Ss* make active attempts to recall or rehearse what they have studied, learning will be more

rapid than if such active recall attempts are not made. Several early studies have shown this. The influence of active recitation is probably motivational in nature, especially if the recall attempts allow knowledge of progress. In the experimental investigation of this variable it is necessary that the time utilized in recitation be added to the study time and that the non-recitation group be allowed the sum of the two periods. In lieu of this, *E* must make allowances for differences in total time spent when he interprets the results.

Sense modality. It is a reasonable question to ask whether *Ss* learn more rapidly when material is presented to the ear or to the eye. The superiority of one method over the other, when found, has not been great and seems to be largely a function of *Ss'* previous experience. If one is accustomed to learn by auditory presentation (as is true for young children), more rapid learning may result than by visual presentation (198, 233). The typical finding, however, is illustrated by a recently published study (134) in which groups of 100 college students were given material to learn by moving-picture presentation, by lectures, and by reading. None of the three methods proved superior to the others in terms of performance measures used.

Amount of material. As the amount of material learned increases, the length of time to learn increases. This is as we would expect. However, as the number of units of material increases, the rate of learning per unit increases (233). Carrying this to its logical extreme we would expect that when the material becomes very lengthy, *S* will be unable to learn within a reasonable period of time.

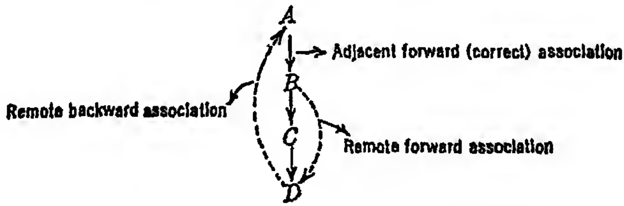
SPECIAL PHENOMENA OF SERIAL LEARNING

Serial learning processes provide us with two related phenomena which in and of themselves have become topics of study. These phenomena are (1) remote associations, and (2) the bowed serial-position curve.

Remote Associations

In the usual serial learning experiment by the memory drum method, *E* writes down all responses made by *S*. It is noted that *S*

typically makes errors in which, let us say, the response which would have been correct if given at the seventh serial position is actually given at the fourth serial position. On the surface there appears to be nothing unusual about such occurrences; errors are common in all learning. These errors which occur in serial learning have special significance, however, depending upon their nature. To schematize the serial learning situation and the kinds of errors which might occur, let *A*, *B*, *C*, *D*, *E*, etc., represent successive serial items. The correct association in the series is *A* to *B*, *B* to *C*, and so on, as represented by the solid lines. If Stimulus *B* elicits



Response *D* (an error) it would be evidence for a remote association in the *forward* direction. If Stimulus *D* elicits Response *A*, it would be evidence for a remote association in the *backward* direction. It is apparent that there can be forward remote and backward remote associations of several different degrees of remoteness, depending on the length of the list. The forward remote association in the diagram has one degree of remoteness; the backward association has two degrees of remoteness if we call an immediate backward association (*D* eliciting *C*) the zero degree of remoteness.

Since errors of all degrees of remoteness, both forward and backward, have been observed, we have a rather complex situation in serial learning wherein each item may be connected to every other item, so that under certain conditions each item might elicit any other item. In the usual situation, the existence of such remote associations is inferred from the errors which occur, but we shall see that remote associations may exist despite lack of evidence based on errors.

Three methods have been used to study these remote associations: (1) the derived-list method; (2) the error method; and (3) the word-association method.

Derived-list method. This method was introduced by Ebbinghaus before the turn of the century. Ebbinghaus has published a vast array of data, all of which were derived from his own learning while acting as both *S* and *E*, yet so thorough and careful was he that his basic results have never been challenged. His measures (N) were the number of lists learned rather than number of *S*s learning a single list as is now the typical procedure.

Ebbinghaus used the derived-list method to demonstrate that remote associations actually exist. To simplify the explanation, let us take the case of a six-item list only, *A*, *B*, *C*, *D*, *E*, and *F*. Initially these items would be learned in the order given. Now, if he wanted to demonstrate the existence of remote forward associations of one degree remoteness, he would construct a second list in which the items were *A*, *C*, *E*, *B*, *D*, and *F* in order. In this derived list, items which had been one degree remote in the original list are now adjacent. According to Ebbinghaus' reasoning, if associations of one degree remoteness had been set up during the learning of the original list, these associations should facilitate the learning of the derived list in which the remote items became contiguous (adjacent forward association). Or, in his words:

> If . . . in the first learning threads of association are spun not merely from each member to its immediate successor but also over intervening members to more distant syllables, there would exist, already formed, certain predispositions for the new series. The syllables now in succession have already been bound together secretly with threads of a certain strength. In the learning of such a series it will be revealed that noticeably less work is required than for the learning of an altogether new series (83, p. 96).

The experimental work to test the hypothesis tends to confirm it, and to establish the principle that the greater the degree of remoteness the less is the saving in learning the derived list.

The error method. This method is merely based upon an analysis of the errors which commonly occur in serial learning. *E* examines each error, determines how far it is out of place, whether it is a forward or backward error, and compiles total frequency of each type. In studying the effects of intra-list similarity we mentioned that the number of errors increased to a maximum at about one-third of the way through learning and

then decreased to zero if the learning was to a criterion of perfection. Those who have used the error method of studying remote associations have, it seems now, determined portions of this error curve.

Word-association method. In using this method *E* requires *S* to learn a serial list to an arbitrary criterion of mastery and then presents the words to him one at a time, with instructions to give the first response which comes to mind. On the association test the words are commonly presented in a different order from that used during learning. In handling the results of the test *E* determines frequency of associations of various degrees of remoteness based on the order of items during learning.

The nature of remote associations. Let us summarize pertinent data concerning remote associations (415).

1. Their frequency decreases as learning increases from moderate to a high degree. Since we assume that these associations did not exist before learning, the frequency must first increase and then decrease as degree of learning goes from low to high.

2. The greatest frequency of remote associations occurs to stimuli from the middle portions of the lists and *are* responses from the middle portions of the list.

3. The frequency of remote associations decreases as the degree of remoteness increases.

It is by now clear that the question is not whether remote associations exist; the data show clearly that they do. The question is whether or not these associations represent associative bonds which have been formed across intervening items in a forward and backward direction. Several theories have been advanced to account for remote associations (233), but all of them are based on assumptions difficult to accept. The facts summarized above tend to support an interpretation which takes the same form as does the interpretation used to account for errors in many other situations: namely, generalization of stimuli. One of the difficulties confronting such an interpretation is the specification of the dimension along which such generalization takes place. The facts given above clearly indicate that the generalization cannot be only along dimensions of similarity such as we have previously discussed. In fact, a very important source of generalization must lie in the serial position factor. Perhaps words at the beginning

and the end of a list have a discriminative feature which those in the middle do not have. The center of the list may be considered more homogeneous than the ends. *S* is more likely to be confused by whether "this word" comes at the seventh or eighth position in a 12-item list than he is as to whether it comes at the first or second position. Any clear-cut interpretation of remote associations in terms of generalization must wait until independent evidence is obtained regarding differential generalization along a restricted serial position dimension.

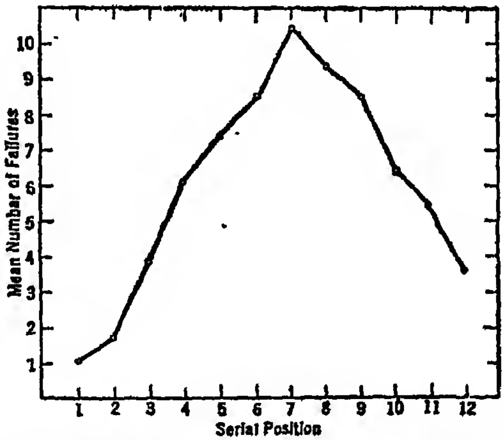


FIG. 71. The bowed serial-position curve. The values are from 64 *Ss* learning lists of 12 nonsense syllables by massed practice at a 2-sec. presentation rate. Data from Hovland (158).

The Bowed Serial-Position Curve

If we plot the mean number of failures for each successive word during mastery of a serial list, we discover the bowed serial-position curve. The curve may be used to determine which words are learned most quickly. The first few items in a list are learned with greatest ease, the last few items are acquired next, and the slowest learning occurs with the item just past the middle of the series, as in the curve for a group of 64 *Ss* in Fig. 71. A very similar curve would be obtained if we plotted the trial on which the word in each position was first anticipated. The curve is sometimes drawn by showing the mean frequency of correct

response made at each position during learning. In such case the curve is inverted.

The serial position curve is not limited to serial verbal learning. It occurs in finger maze learning to a certain extent, especially if *S* is given a set for speed. Recently, the curve has been reported in punchboard maze learning (185). There are also many incidental observations, such as memorizing bus stops, which support the generality of the curve.

The impression that the bowed serial-position curve will not change with changed conditions must not be given. If the seventh item in a ten-item list is typed in red ink and all others in black, the seventh item will be learned more rapidly than either the sixth or eighth item, thus changing the shape of the curve (184). There is, of course, nothing magical about red ink—the same result would have been obtained had the seventh item been in black and all others in red. McGeoch (233) reports an unpublished study by McGurty in which ten-item lists of nonsense syllables were constructed so that the formal similarity of the first and last three items was very high, and the middle four items quite dissimilar. Under these conditions the bow disappears and the items in the middle are learned almost as rapidly as the first and more rapidly than the last three items.

Yet, in the normal course of learning serial tasks, the bowed curve is a very persistent phenomenon, and when the structure of the series is not markedly altered, it will usually be found. There are common sense arguments which might suggest that the bowed curve is an artifact of the situation. Rehearsing items between trials is often suggested as an explanation by students who come upon the curve for the first time. Yet learning in which rehearsing is prevented or in which there is no rest between trials only slightly alters the shape of the curve.

The second argument which is usually advanced by students is that the first and last items are learned most readily because they *are* first and last. Such an explanation is not acceptable until it is shown by what means first and last gain a differential status and how such a status could influence learning.

As in the case with remote associations, there is no generally accepted account of mechanisms which produce the bowed serial-position curve. Very likely when one is explained both will be

explained, since the relationship between the two is seemingly clear.

MAZE LEARNING

A tremendous volume of data is available on maze learning, especially rat maze learning. In our discussion of variables influencing rate of learning, we included maze studies as illustrations. In addition to the general principles which cut across multiple response learning as a whole, certain phenomena are largely peculiar to the maze situation as such. More correctly, perhaps, they are intrinsic to learning tasks which require locomotion through space. Very briefly we shall consider these general phenomena for human and animal maze learning.

Human Maze Learning

Learning method. One of the outstanding facts of human maze learning is the lack of consistency among *S*s concerning the methods used in learning a multiple-U or multiple-T finger or stylus maze (422). Maze learning is, after all, a task which does not have many counterparts in real life situations. *S*s do not have much background experience to bring to the situation. After preliminary practice *S* will usually hit upon a procedure which becomes his characteristic method. Husband's work (176, 177, 178) has shown that a majority of *S*s will discover the *verbal method* in which they memorize the turns in successive order as left, right, right, left, and so forth. Other *S*s report that they use a more highly motor method in which the *kinesthetic cues* of distance and "feel" play a predominant rôle. Still others report what has come to be known as the *visual imagery* type of learning in which *S* translates the kinesthetic cues of distance and proportion into a visual image of the maze. There are in addition, combinations of the methods.

The verbal method usually brings about the most rapid acquisition.* A puzzling problem is why more *S*s do not immediately use the verbal method, since most problems they met in school and elsewhere require solutions which are verbal in nature. The

* In some class-room experiments conducted by the writer, *S*s instructed to use the verbal method learned multiple-T finger mazes much more rapidly than did *S*s left to their own devices.

spatial factor present in the maze may provide difference enough to prevent the immediate initiation of verbal solutions, although this is largely speculative at present.

Orientation. Although it is difficult to get systematic data on the matter, we may be quite sure that the process of maze learning involves spatial orientation which is roughly equivalent to *S*'s saying: "I know the end of the maze is in that direction." We do not know how precise this orientation is nor how many errors can be attributed to *S*'s attempts to go directly toward the goal rather than in a circuitous route as the true path may actually be.

Animal Maze Learning

The goal-gradient. In human maze learning there is a tendency for the correct alleys near the beginning and end of the maze to be mastered more quickly than those in the middle. This order of acquisition is less clear in maze learning than in serial verbal learning but is usually found if the maze is constructed so that considerable homogeneity previals within the maze and if *S* is set for speed. In animal learning, this order of learning specific correct responses is not typically found. Instead, a rat will eliminate the errors (entering blind alleys) nearest the food box first and then those successively in a backward direction so that the greatest number of errors are made in the blinds immediately outside the starting box. Thus, the elimination of the errors is in a backward direction. An explanation for this has been given by Hull (163), and the phenomenon has come to be known as the *goal-gradient.* The goal-gradient may be obscured and modified by other factors as recent work of Hull and his students has shown (e.g., 166, 366), but it does, nevertheless, stand as a fairly general principle of animal maze learning.

Orientation. Part of what a rat learns in a maze appears to be an orientation toward the goal. It is as if, after a few trials, the animal learns that the goal box is "over there." We do not need to say the rat learns that he has to go north by northeast to get to the goal; there are probably many stimuli, both internal and external, which become conditioned to running toward the food box, thereby serving as orienting cues.

There are controversial theories regarding the nature of orientation and its importance for learning. In our chapter on theory

we shall return to the contentions, along with a more careful consideration of just what *is* learned by a rat in the maze.

Sense modalities. Attempts to determine the sensory modality used by rats in learning a maze have taken the form of eliminating certain modalities by minor operations. Honzik's work (155) is an illustration. Various groups of rats had various sense modalities destroyed, that is, vision was destroyed in one group, audition in another, the sense of smell in another, and various combinations of modalities in still other groups. In general, experiments of this kind show that some retardation in learning results from the destruction of any modality, and the greater the number of sense modalities destroyed the slower is the learning.

CONCLUDING CONSIDERATIONS

In conclusion we should point to a very difficult problem which arises if we wish to determine the conditions which bring about *optimal learning efficiency* for any given task for a defined population. We have seen how optimal learning efficiency may be determined for a *single variable* with other conditions held constant. But when we wish to consider optimal learning conditions for all variables combined the problem is quite complex, even for a simple task.

The student who is primarily interested in the practical application of research results must be re-educated if he holds any delusion that by a few experiments the optimal conditions for learning a task can be determined. Learining efficiency may be *increased* by judicious application of principles derived from a few experiments, but a conclusion that maximum learning efficiency has been attained represents wishful thinking and far outruns proved fact. The optimal methods of presenting a learning task will be learned only by a long series of researches in which the interaction of very many variables is determined.

During World War II a great deal of research work centered around attempts to improve methods of code learning. Quite a number of these analytical researches by Keller and his associates are described in print although probably many others have not been published. These researches show that much progress was made toward increasing code-learning effectiveness, but these *E*s

would surely be the first to suggest that the conditions of maximal learning efficiency have not as yet been unearthed in spite of the series of systematic studies. So many variables and combinations of variables are possible that many more researches will be needed.

When we keep in mind the multitude of serial or multiple-response tasks which are performed day after day by people in the process of earning a living, e.g., tasks such as typewriting, comptometer operating, typesetting, and assembly tasks, we can scarcely overestimate the need for research which will result in sound data leading toward a realization of optimal learning conditions. At present we can only caution ourselves against hastening to a conclusion that bits of knowledge can be pieced together easily into a total pattern defining the conditions for optimal learning efficiency.

SUMMARY

1. In this chapter we have evaluated experimental methods and results in learning tasks which require the integration of a series of responses to a series of stimuli. It is this form of learning by which many skills required by formal education are acquired.

2. Typical laboratory tasks were described which fall along a dimension of degree of motor involvement. This dimension extended from verbal learning, representing low motor involvement, to pursuit rotor learning, representing high motor involvement.

3. In the discussion of response measurements we were concerned largely with trials-to-criteria method of plotting curves.

4. Variables influencing rate of learning were as follows:

- *a.* Massed vs. distributed training
- *b.* Type of material
  - (1) Intra-task similarity
  - (2) Meaningfulness
  - (3) Affectivity
- *c.* Knowledge of performance
- *d.* Miscellaneous

5. Two special phenomena of serial learning, remote associations and the bowed serial-position curve, were reviewed.

6. A very brief analysis of maze learning was included.

7. Finally, we considered the problems associated with determining conditions which bring about optimal learning efficiency.

CHAPTER XIII

Learning III — Thinking

INTRODUCTION

Other titles might well have been given to this chapter, such as *Reasoning*, *Problem Solving*, or *Creative Thinking*. A procedure which one psychologist may say demonstrates thinking may be said by another to demonstrate reasoning; still another may say it should be called problem solving. We need not be too concerned about the label as long as we indicate the descriptive relationships between the material included under these various terms and the material covered in the two previous chapters. Few will fail to recognize that the present chapter covers the chief area of behavior by which man traditionally sets himself apart from other animals. For surely, here are the skills with which airplanes are designed, sonnets are written, and atoms harnessed. Here indeed, some may say, should be the starting and stopping point of psychological investigation. The records of rats jumping up and down on a lever, or a finger pushing its way along a maze may pale into insignificance before the contemplation of the behavior which rears the achievements defining the advance of civilization. But, let us get along with the analysis and forget the chest-thumping.

Descriptive characteristics of thinking. Over-all, we have been using the descriptive dimension *complexity of task* to systematize our study of learning. In the conditioned response situation it was found that a single response became associated with a rather discrete stimulus; the multiple response situation required that *S* integrate a series of responses to a series of stimuli. The thinking situation is still more complex in that *S* is required to determine the relationships among various stimuli and the "correct" response is usually forthcoming only when these relationships have

been perceived. This description does not, of course, place thinking at a discrete point on the dimension. We shall find that tasks which have been used to study thinking form a "band" or "zone" along the dimension of complexity, the lower portion of the band shading into multiple response learning.

The diversity of the tasks which have been used in studying thinking makes complete description in terms of a single dimension rather difficult. We should identify other characteristics which are sometimes needed to demonstrate the relationships between thinking and the learning situations discussed in previous chapters.

1. The thinking situation places great emphasis on the discovery of the correct response. The "search" (182) is an integral part of thinking behavior. In terms of discovery or search, thinking is much like Conditioning Procedure II (Skinner-box), but is quite unlike rote learning in which discovery of the correct response is at a minimum. Although the discovery process may be facilitated because *S* eliminates certain possible responses by "logical" considerations, we should recognize that logical procedures are in themselves a subject of study.

2. In multiple-response and conditioned-response learning situations the fixation process is basic, whereas in some thinking situations it is of secondary importance. The correct response in some thinking situations is a single unitary principle, and once it is discovered learning may be fairly complete.

3. The thinking situation places heavy emphasis on symbolic representation and manipulation. Overt behavior is at a minimum, and when present is at best only a way of testing or verifying the validity of symbolic manipulation.

It will probably be clear that these descriptions are simplified and quite crude. Unfortunately, no psychologist has attempted a systematic scaling of tasks along such descriptive dimensions. Yet, if we can perceive various relationships among tasks and processes needed to learn these tasks, it may prevent us from hastily concluding that conditioning is one "type" of learning, multiple-response another, and thinking still a third. The point of view advanced here is that the terms (conditioning, multiple-response learning, thinking) simply refer to modal points of laboratory operations and may be relatively inconsequential in

the formulation of basic principles of learning. We do not assume that there are different *kinds* of learning; the relationships among all tasks used to study learning are too apparent to admit such a distinction *at this time*. It is a matter for the future to produce the evidence requiring abandonment of this position and establishing the fact that certain distinctions are necessary. Later we shall point out that some psychologists hold an opposite point of view and insist that there is warrant for making those distinctions now.

Status of research on thinking. Upon the completion of this chapter we shall realize that, compared with other areas of learning, the experimental study of thinking is somewhat retarded. Sound empirical facts concerning thinking are relatively few in number. The experimental method has been used only haphazardly in studying thinking, and there has been little continuity in the tasks from experiment to experiment. We shall find that when we attempt to draw up a list of variables which cause the rate or efficiency of thinking to vary, some improvisation will be required. Probably the most telling revelation of the current status of research on thinking is the fact that during the last five years less than a score of truly experimental studies have been published in standard journals. Indeed, if we were to allot space in this book on the basis of frequency of recent research on a topic, we should relegate thinking to a few pages tacked on to some other chapter. However, in the belief that thinking is a topic to which experimental psychology can and must contribute, we shall attempt to treat the topic in much the same fashion as the previous two topics on learning.

We may ask *why* research on thinking is retarded. There are probably several reasons.

1. It appears that many psychologists have felt that the process of thinking, being covert in nature, is somewhat beyond the reach of the experimental methods which we have discussed thus far. According to this point of view, thinking takes place in a closed system and can be reached only by a form of introspection. In applying the introspective method to thinking, *S* is given a problem and while attempting solution, or after arriving at a solution, he describes the conscious content of his experience, he reports all thoughts, ideas, reasons for doing this, for doing that, and so forth. This, be it noted, is not the precise and sophisticated

form of introspection used in the early years of psychology to study chiefly more purely sensory experiences.

Such an approach to the psychology of thinking has little defense if it be the central aim of an investigation. Experimentation concerning thinking can and should be carried on primarily in exactly the same fashion as experimentation on animals. A rat does not sit down at the end of a maze and explain to *E* how it happened to take the right path instead of the left path. No one is forced to adopt introspection as his primary method of investigation just because the processes needing investigation appear covert. While careful introspective reports will yield some valuable data and while introspection constitutes a perfectly valid method of research, there is no reason why the usual experimental methods cannot be applied to the study of thinking; that is, we may systematically vary the stimulus conditions and observe the response changes. The fact that we *do* know more about how a rat solves a maze than we do about how a student solves a geometry problem only adds to the confidence with which we may attempt to apply straightforward experimental methods to the processes of thinking.

2. The failure to get on with the experimental study of thinking may also reflect the procedural bias of many psychologists who prefer to start at the simpler level of behavior and work toward the more complex. According to them we are not yet ready to study thought processes. As the simpler situations become more thoroughly analyzed, perhaps research workers will finally turn their attention to the complex situations.

3. A third explanation for the status of research on thinking has been indicated by Heidbreder (145) when she suggests that in most cases there has been a failure to ask effective questions about the thought processes. By effective questions she means those which are provoked by a theory, and those questions, it may be added, usually refer to the relationships between manipulable environmental conditions and changes in performance.

Plan of chapter. Response measures used for assessing performance on thinking tasks are usually *time* to solve, *trials* to solve, or *errors* made. They are handled in much the same way as the response measures of multiple-response learning, so we shall not need to discuss them separately.

The remainder of the chapter is divided into three general sections:

1. Problems illustrating thinking
2. Variables influencing rate of thinking
3. Miscellaneous problems

PROBLEMS ILLUSTRATING THINKING

Although we do not have thinking problems which may rightfully be called standardized laboratory tasks, in the sense that we call Skinner-boxes and nonsense syllables standardized, we should outline the general nature of the area of thinking by discussing briefly certain tasks which have been used.

Rational learning. Rational learning problems have not been used in many experiments during recent years. Yet we shall discuss the procedure in detail because it typifies the least complex of thinking tasks, and therefore will give us an anchor point for evaluating the more complex tasks. Rational learning provides a ready transition from multiple-response learning to thinking. Peterson's problem (290) is probably the best known. Typically, ten letters, A, B. . .J are paired with numbers one through ten in a random fashion. *S* is told that each letter has a number from one to ten assigned to it and that no number is assigned more than once. *S*'s task is to discover and remember the number which goes with each letter, so that when a letter is presented as a stimulus he can give the correct number assigned to it.

When *E* gives the first letter, A, on the first trial, *S*'s responses are pure guesses from among the ten responses available. After each guess *E* reports "right" or "wrong" and *S* continues guessing until he arrives at the correct number for A. *E* then presents B as the second stimulus. Now note, that for stimulus B the number of possible correct responses has been reduced by one, since the correct number for A cannot be correct for any other letter. If four had been found correct for A it would be "irrational" for *S* to guess four for B or for any other letter. For each letter in order, then, the number of possible correct responses is reduced by one. To the last stimulus, J, there can be only one number as the response, and if *S* remembers all the correct responses for the previous nine letters he should get the correct response to J on the

first try. Usually the series is repeated until *S* is able to give all responses without error. Also, *E* varies the order of the letters from trial to trial. If he does not do this, the problem becomes a serial learning task, since *S* will ignore the stimuli and merely learn the sequence of responses.

The rational learning problem has two characteristics which we should keep in mind. First, we note the heavy emphasis on immediate memory. If *S* can remember the correct responses for previous letters his task is greatly simplified. Rational learning in this respect depends upon good memory. Because of this important emphasis on immediate memory, we may expect that factors which influence memory will effect rate of rational learning. Secondly, the rational learning problem is very similar to paired-associate learning, differing only in the fact that *S* has to discover the correct response from among a rather limited number of possible responses. The acquisition curve of a rational learning problem (usually an error-reduction curve) is much like the usual multiple-response learning curve.

Concept formation. Concept formation is the most common laboratory task used to study thinking, even though little standardization of concept materials has been attained. Much of our data on thinking has its origin in concept formation, which is a kind of inductive reasoning or thinking. "What is given in the inductive problem consists of specimens, and the result to be attained is a definition, or at least a working knowledge, of the class represented by the given specimens" (422, p. 801). In the laboratory *S* "must develop an effective response to a class of objects and a different response to objects not belonging to this class" (422, p. 801).

The formation of language concepts is an integral part of one's education. Dogs become distinct from cats even though both have four legs, and even though both are known as animals. Concept formation is a generalizing process which follows the discrimination of similarities and differences among objects. Botanical and zoölogical classifications are based upon recognized differences and similarities among various specimens. The greater the number of specimens included under a single concept, e.g., fishes, the greater is the number of specimens which have a common descriptive dimension.

Studies of concept formation have taken two forms.

1. There are studies in which *S* must learn a new name for an old concept, when the major problem is to discover the concept. For example, from the following problems can you tell what concept is now named *JUNKO?*

NO JUNKO:	$4/2 = 2$
JUNKO:	$1 - 1 = 0$
NO JUNKO:	$5 \times 1 = 5$
JUNKO:	$18 - 3 + 8 = 23$
NO JUNKO:	$9 + 3 + 4/2 = 14$
JUNKO:	$6 \times 2 - 4 = 8$

You will find that only problems which call for *subtraction* are *JUNKO;* hence, *JUNKO* is a new name applied to the concept of subtraction.

2. There is another small group of studies in which *S* has to learn a new concept. The first time you learned the meaning of subtraction as opposed, let us say, to addition, you learned a new concept.

The task of evolving concepts usually consists of setting up successive hypotheses and testing those hypotheses. One particular kind of concept-formation problem is especially useful in demonstrating this. Sometimes called the multiple-choice problem, this one may be illustrated with numbers. *E* presents *S* with successive series of numbers and informs him that in each series one number is correct. *S*'s task is to discover which number is correct and to develop a principle by which we can choose the correct response on each successive series. Let us say the first series is 3, 4, 5, 6, 7 and *S* discovers that 4 is the correct response. The second series, 9, 10, 11, is presented. What did *S* learn on the first series that might aid him in selecting a number from the second series? He remembers that 4 was correct on the first series; 4 is an even number. For want of any other hypothesis, he chooses 10 on the second series because it too is an even number. *E* informs him that 10 is incorrect. Subsequently, *S* discovers that 9 is the correct response on the second series. Then he must ask himself what relation 4 in the first series has to 9 in the second series. A third series is presented, 1, 2, 3, 4, 5, 6, 7, and he finds that 3 is the correct response.

S continues to test hypotheses and sooner or later will discover

that the correct response is always the number which is one left of center; thus, one-left-of-center is the concept to be applied. *E*, of course, may make the correct concept a very difficult one, such as alternating between one left of center and one right of center. The discovery of the correct concept is greatly facilitated if *S* is allowed to keep all previous series before him with the correct responses marked. If these are not present so that *S* must rely on memory for previous correct responses, many incorrect hypotheses will be tried because *S* cannot remember correct responses for some of the series he has already had.

Miscellaneous. Many tasks of the parlor game variety have been used in studying thinking. Mechanical puzzles, jigsaw puzzles, verbal conundrums, mathematical problems, and so forth, all have been used at various times, but with insufficient frequency to warrant specific descriptions.

VARIABLES INFLUENCING RATE OF THINKING

When we refer to variables influencing rate of thinking we mean variables which determine the time, trials, or errors to solve a problem called a thinking problem. The criterion of solution is set up by *E* and will, of course, vary with different problems. Thus, just as we studied variables determining rate of conditioning and rate of multiple-response learning, we shall also study variables determining rate of thinking.

Massed vs. Distributed Training

In conditioning and multiple-response learning, distributed training usually brings about more rapid learning than does massed training. However, the available data indicate that the generalization as thus stated does not necessarily hold for thinking problems. We shall evaluate three studies.

Cook (66) used problems somewhat similar to jigsaw puzzles. One was called a T-Puzzle, the other a Cross-Puzzle (Fig. 72). Putting the plywood pieces together was a more difficult job than it may appear to be. Piecing a puzzle together once constituted a trial, and the basic response measure was time per trial. Massed training was defined as 5 seconds between trials, and distributed training as 24 hours between trials. *S*s were given 15

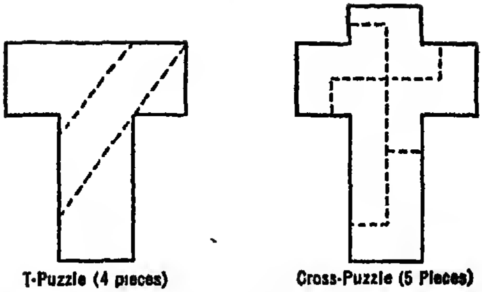


FIG. 72. Puzzles used by Cook (66). The dotted lines show how each puzzle was cut.

trials on the Cross-Puzzle and 20 trials on the T-Puzzle; the 20 *Ss* used worked on both puzzles by Design Method III.

The results for the series of trials on each puzzle are shown in Fig. 73. The first 3 trials have been omitted because of the great amount of time required to solve initially these particular puzzles. In addition, successive blocks of 3 trials each have been com-

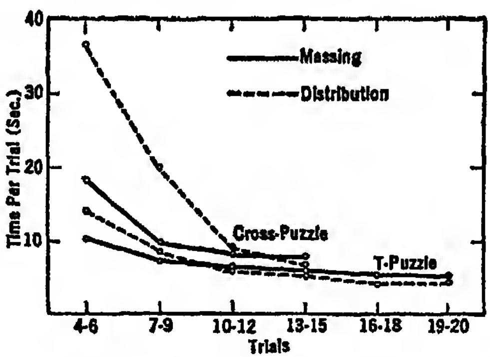


FIG. 73. The effect of massing and distribution of trials on puzzle-solving. Data from Cook (66).

bined. The results are quite clear-cut; massing is superior early in learning, especially in learning the Cross-Puzzle. With successive trials the distributed training results in performance which is just as good as that achieved through massing and there is some indication that performance under distribution eventually becomes somewhat, though insignificantly, better than that under

the rival method. The one perfectly clear finding, however, is that early in the learning of a puzzle problem the massing of training results in more rapid learning than does distribution of training.

In a series of experiments Garrett (103) used one task which might be considered as belonging in the field of thinking. *S* was told to construct an *artificial* language, using a series of words and a series of rules about those words. The rules indicated how to change nouns from singular to plural, how to indicate past tense, and so forth. These rules, of course, were new ones in the sense that they were not the rules by which our own or any other existing language is written. After providing illustrations of the proper use of the artificial language, *E* gave *S* 12 sentences which were to be translated into the artificial language. Massed practice consisted of translation periods following each other immediately, and distributed practice was defined by one translation period every 3 days. The practice periods in one experiment were 2 minutes in length and in another 4 minutes.

The results tended to favor massing of practice in terms of amount of gain made from one practice period to the next, although the differences between performance under massing and distribution were not great. A pertinent fact is, however, that in two other conditions Garrett used simpler tasks and the results tended to show that distribution produced the most efficient learning—the same as the typical finding discussed in previous chapters.

An experiment by Erickson (90) utilized a puzzle, the nature of which need not concern us, except to say that solving it provided a difficult task. The distributed practice group was given 5 minutes of practice every 48 hours, and the massed group, only one period of practice, until an arbitrary criterion of solution speed was attained. In terms of trials or time to reach the speed criterion, the performance of the massed group was much accelerated over the performance of the distributed practice group, being 63 per cent superior in terms of trials and 44 per cent in terms of errors.

The results of these three experiments agree very well; with these complex learning tasks in which *S* must grasp and retain relationships among various stimuli in order to respond

appropriately to those stimuli, distribution of practice is not superior, and there is clear-cut evidence in two of the three studies that massing is superior at certain stages of learning.

Let us note, however, that limitations must be placed on any generalization concerning massed and distributed practice in thinking. In no instance did *E* explore a dimension of degree of distribution. The smallest interval of distribution was 24 hours and the interval between massed trials was a few seconds. It is still quite possible that, let us say, an hour between trials, would bring more rapid learning than would 5 seconds between trials. In short, our generalization on the influence of massing and distribution on thinking is quite restricted. The degree of distribution (with various length work periods) has not been adequately explored.

Type of Material

From our own experiences we recognize that problems involving thinking vary in difficulty. It is the purpose of this section to attempt to identify characteristics of material which produce differences in difficulty and therefore differences in rate of thinking. Unfortunately we do not have a great deal of data with which to work. Very few experiments have been concerned directly with this problem and most of them are experiments in concept formation. Consequently, any generalizations at which we may arrive must be considered to be quite tentative.

We may identify three variations of material which influence the rate of thinking: (1) length of task, (2) nature of concepts, and (3) intra-task similarity.

Length of task. The length of a given task and difficulty of learning are, of course, positively correlated. We will consider briefly two studies which investigate this relationship.

1. *Puzzle Solving.* The disc transfer (or pyramid) puzzle, was used in a study by Cook (67) to show the relationship between length of task and difficulty. The pyramid puzzle in simple form consists of three "stations" which we may call A, B, and C. *E* places two discs at A, the smaller on top of the larger. *S* is instructed to move the discs from A to C, moving one disc at a time and never placing a larger disc on top of a smaller one. Station B is used as a "way" station. With two discs a minimum

of 3 moves is required, with three discs, a minimum of 7 moves, and with four discs, 15 moves.

Cook matched three groups of *S*s on the basis of an intelligence test before giving one group the two-disc problem. *S* worked on the problem until he could solve it in the minimum number of moves. The mean *total* moves required for the two-disc problem was 3.5, for the three-disc problem 24.3, and for the four-disc problem, 182.7. It is clear from these data that the number of moves actually required increased at a much more rapid rate than the minimum number of moves theoretically required to solve the problems.

In a second experiment Cook used a modified form of Design Method III. One group of *S*s solved all three problems in order, from the two-disc problem through the four-disc problem, whereas another group solved the four-disc problem first, followed by the three- and two-disc problems. The group solving the simple to complex problems showed very little transfer from one problem to the next, whereas the group solving the complex to simple problems showed large transfer effects. Here then is a task in which we have differential transfer effects in going from one condition to another, depending on the order of conditions, and we must conclude, as Cook did, that counterbalancing should not be used for experiments on this task (cf. Chapter X).

2. *Concept Formation.* An experiment by Reed (304), while not one in which length of problem was varied directly, is related to the problem of length and time to learn. Reed's *S*s were presented with a series of 42 cards. On each card were 4 words. From the 42 cards *S* was to learn 6 concepts; hence, each concept was represented among the words on 7 cards. The names of the 6 concepts were nonsense syllables. To take an illustration; the syllable *KUN* was the name for a concept represented among each of the following 7 groups of words, each group of words printed on a separate card:

1.	horn	leaf	monkey	debt
2.	fame	ought	tiger	saucer
3.	line	people	elephant	sound
4.	uncle	friend	sheep	pear
5.	crowd	sail	deer	string
6.	horse	circle	paid	scholar
7.	carrying	died	cow	ruler

In the 7 groups only the concept of an animal is common to them all and since animal names appeared with none of the other syllables used, *S* eventually discovered that *KUN* was the response to a card on which an animal name appeared. Since five other concepts were to be learned, each represented by 7 cards, the total of 42 cards is accounted for.

E showed each card for 3 seconds and then pronounced the syllable which went with the card. On successive trials *S* named it correctly, *E* said "correct"; if *S* failed to say the syllable or said the wrong syllable, *E* pronounced the correct syllable at the end of the 3-second period. The index of rate of learning was the number of promptings needed before *S* correctly identified all cards by calling out the appropriate syllable within the time limit allowed. The order of the cards was varied from trial to trial, and cards representing the same concept never followed one another.

With one group of *Ss* Reed had 4 words on each card as indicated above, but with another group 8 words were presented on each card although the time of presentation of each card was the same. The question asked was how much difference in the rate of learning resulted from differences in the number of names on each card even though the number of concepts to be identified was the same.

The results of this experiment show that the most rapid rate of learning took place when only 4 words were on the cards. On the average, the number of promptings to learn the 6 concepts with 4 names per card was 30.7, whereas with 8 words per card the mean was 40.0. If we assume that the intelligence test used is a valid matching instrument and that the two groups were therefore not significantly different in initial ability, we may conclude that the extra names on the cards made the task more difficult. The greater the number of words the greater the number of possible hypotheses which *S* had to eliminate before formulating the correct one.

The nature of the concept. Two series of studies afford data on rate of concept formation as a function of the nature of the concept to be attained. We shall take an appropriate study from each of these series. These studies will clarify what is meant by "nature of the concept."

1. *The Welch-Long Studies.* The subject of these studies is concept formation in children. One study was concerned with concept formation as a function of "level of abstraction," a phrase to be defined by the experimental operations.

The situation Long and Welch (222) used in studying this problem is represented in Fig. 74. A, B, and C stand for objects. When A and B are together, a light is on; when B and C are together no light is on; when A and C are together the light is on again. On test trials *S* is asked which of the three objects, A, B, or C, will make the light go on. A, since it is common to both situations in which the light is on, is the correct answer. With children as *Ss* pictures of objects are inserted for the letters. The first pair, let us say, a lion and a camel, are presented, and *E* points out to *S* that the light is on. Then, the second pair, a horse and a camel are presented, and *E* is careful to show the child that the light is not on. In the third series the lion and the horse are presented, and again *E* points out that the light is on. On the test trial *S* is presented with pictures of a lion, horse, and camel and is asked to choose the one which will make the light go on.

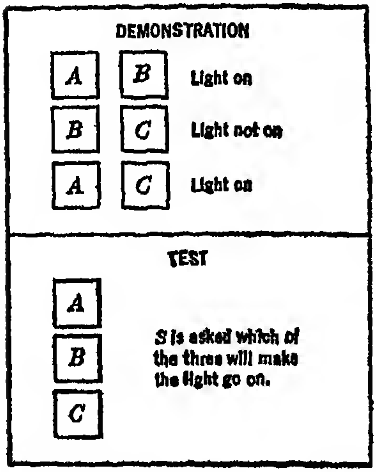


Fig. 74. Illustration of stituation used by Long and Welch (222) in their studies of concept formation.

The above illustration indicates concept formation at the object level. "Light on" becomes a new name for an old concept of lion. Now, let us move up a level and present *S* with the following series of pictures:

lion	— pencil:	Light on
boy	— bus:	Light not on
camel	— carrot:	Light on

On the test trial we present *S* with three pictures, a horse, a battleship, and a bean and ask *S* which one will make the light go on. If he has "caught on" he will pick the horse, since the light

was on in each case where a sub-human animal picture was present. This, then, is a first level abstraction.

For the second level abstraction, we present *S* with the following series of pictures:

man — auto:	Light on
pear — rowboat:	Light not on
lion — corn:	Light on

Three test pictures, bus, bean, and camel, are then presented. The correct response is *camel*, since the light has been on for lion and man, and both are animals.

Tests at the various levels of abstraction were given to 40 children, 8 to 11 years of age, and a great many different objects were pictured. Furthermore, the positions of the right and wrong pictures were randomly changed from trial to trial so that choices would not be based on position cues. From their tests Long and Welch conclude that concepts at the abstract level were more difficult to learn than those at the object level. However, once *S* discovered a concept at an abstract level he was able quickly to master additional problems at that level.

2. *The Heidbreder Studies.* Heidbreder planned a series of researches to discover under what conditions *Ss* change their concepts. In order to study this problem Heidbreder wanted to use concepts of known difficulty. Since no such standardized materials were available, initial experiments were set up to derive empirically concepts of known difficulty. An unexpected outcome of these initial experiments was the finding that certain classes of concepts were formed faster than others. Concepts of concrete objects were formed most rapidly, those of spatial forms next most rapidly, and number concepts were attained last (144). Let us look briefly at the method used to discover these facts.

In most of the experiments nine concepts were learned, three each of the above classes. The material was presented pictorially on a memory drum, and *S* was asked to associate a nonsense syllable with each concept. Thus, the concept name for faces was *relk*. Each picture appeared for 4 seconds, and after 2 seconds of this period *E* pronounced the nonsense syllable, thus indicating the correct name. There was 16 series, each being made up of nine pictures, one corresponding to each of the nine concepts. For

example, there were 16 faces, each face different, 16 trees, each tree different, and so on. *E* presented each series in succession until *S* gave two successive perfect recitations.

The three object concepts were *face*, *building*, and *tree*, and as indicated above, were represented in each series by actual drawings of faces, buildings, and trees. One of the three spatial forms was a circle, but in the picture series a plain circle was not presented. Rather, a circular wreath represented the circle concept on one series; a wagon wheel on another, a ball on still another, and so forth. A second spatial form looked like an equal sign with a cross-hatch. Like the circle, it was presented in several different ways, although in each case the basic form was there. A third form appeared as a small circle with two antennae.

The number concepts were presented as objects. Thus, for the concept *three*, one series might have shown three nails, the next, three houses, the next, three butterflies, and so forth. *S* was to discover that the common characteristic of all these representations was "threeness."

Heidbreder's results show that the object concepts were learned to mastery by 63 *S*s in an average of 3.6 trials. The spatial forms required an average of 4.9 trials, and the number concepts, 8.4 trials. These differences were highly significant statistically.

Although Heidbreder has demonstrated that such an order of concept attainment is not inevitable (146), in general, the order has proved surprisingly resistant to changed conditions. Because of definitional difficulties, Heidbreder prefers not to use the term *abstract* to indicate the obtained order, saying, for example that it is based on successive levels of abstraction, or that the more abstract the concept the slower is the rate of learning. In the Long-Welch study, it is true, we did refer to successive hierarchies of classification as levels of abstraction. Although such a hierarchy is not present in the case of Heidbreder's concepts, it is, nevertheless, approximately true that as we go from objects to forms to numbers we get more and more removed from "thing-like" characteristics. Yet, even if this is true, and even though we might speak of levels of abstraction, we still have to show how the characteristics which define levels of abstraction produce varying learning rates. This cannot be done satisfactorily at the present stage of research.

In the future, level of abstraction may be shown to be a special case of meaningfulness as we have used the term *meaningfulness* in the preceding chapter. Some evidence suggests this. Wilkins (411), working with syllogisms, found that level of difficulty was directly related to level of familiarity. A syllogism is a form of deductive thinking wherein *S* is given two or more premises and is asked what conclusion can be reached. In concept formation we go from the specific to the general, whereas in syllogisms we go from the general to the specific. We may illustrate the syllogism by continuing to overwork Socrates' mortality:

> All men are mortal;
> Socrates is a man;
> Therefore, Socrates is mortal.

Wilkins gave syllogisms stated in familiar terms and syllogisms stated in formal terms, as illustrated here:

Familiar	*Formal*
Some students are bright;	Some x's are y's;
George is a student;	Some z's are x's;
Therefore:	Therefore:

The syllogisms stated in familiar, as opposed to formal terms, were most frequently solved correctly. If we considered the essential components (students, bright, George, x, y, z) of such syllogisms it is quite likely that we could demonstrate different degrees of meaningfulness. Would this be the same as levels of abstraction?

Intra-task similarity. We have seen that this factor is very important in determining the rate at which multiple-response learning takes place. Most thinking problems involve the perception, implicit manipulation, and short-term retention of relationships among stimuli. In so far as perception and retention are based on similarity of stimuli, this variable should also determine the rate of thinking. Similarities among various aspects of a problem may be so confusing that *S* may be unable to grasp the problem quickly. For example, Johnson gives a problem which most people will have to read more than once in order to understand:

> "The ages of a man and his wife are together 98. He is twice as old as she was when he was the age she is today. What are their ages?" (182, p. 206).

The probable importance of similarity in grasping a sequence of thoughts is clearly demonstrated on a well-known weekly radio quiz program. Each week a participant is asked to repeat a sentence after hearing it one time only. The sentence, using the same pattern each week, runs about as follows:

"Ted is in bed," said Jed to Ned; "Ted is in bed, Jed," to him said Ned.

The fact that a relatively small percentage of the participants is able to repeat such a sentence is probably caused by the confusing similarity relationships within the sentence.

At the experimental level, however, we have very little evidence on the rôle played by intra-task similarity in the thinking process. Welch (403) has made an analysis of the results of experiments he has performed with Long and has concluded that similarity is an extremely important factor in determining concept formation. It should be clear, however, that concept formation is a well-practiced task and may not be considered a very complex type of thinking. Certainly we are in no position at present to generalize to other tasks from principles based on concept formation. Nevertheless, Welch's analysis appears to have good applicability to concept formation and it will be worthwhile to outline the essential points of it.

Concept formation studies such as Welch has conducted involve the use of positive instances of the concept and negative instances of the concept, the latter being illustrations of what the concept is not. According to Welch, two principles of similarity are important in such studies. (1) The greater the similarity between the positive and negative stimuli the slower is the rate of concept formation. (2) The greater the similarity between the positive stimuli the faster is the rate of concept formation.

Both of these principles have common sense appeal, and both may be derived readily from studies of transfer of training and discrimination learning. The first principle is demonstrated in studies in which an animal must learn to respond, let us say, to a circle, and to make no response to an ellipse. The second principle is shown by any study of transfer in which the same response had to be acquired to similar stimuli. We cannot, of course, subsume all known facts about concept formation under these principles,

as Heidbreder's results demonstrate, but certainly the principles may form a part of the general theoretical structure which will be needed to explain complex concept formation.

Set

In our discussion of set in Chapter VI we indicated that there are two ways by which *E* may influence *S*'s set: (1) by instructions, and (2) by controlling *S*'s past experiences. We have identified the directive function of motivation with set; set determines the stimuli to which *S* attends and therefore will influence his responses in a problem situation. *E*, by appropriate instructions, may specify the stimuli to which *S* should attend and may also specify the mode of attack which is likely to be most fruitful. Instructions, in a sense, provide the same localization of behavior in humans as food smeared on the bar in the Skinner-box for rats. The second method of controlling set is by giving *S* preliminary training in a certain mode of solving problems and measuring the influence of that training in a new but similar problem situation. This method, in effect, provides a continuation of our study of transfer of training.

The Maier experiments. Much of what we know about the influence of verbal instructions on problem solving has been obtained from a series of systematic experiments by Maier. Maier assumes that thinking involves primarily the combination of past experiences to bring about perception of a new relationship among stimuli. It is this new perceptual relationship which provides the solution to a problem. One factor which Maier believes to be important in bringing about this new relationship is what he calls *direction*. The concept is the same, or very similar to what we mean by *set.**

1. *Experiment I.* The first experiment (236) was designed to determine the effectiveness of recall of past experience in problem solving. *S*s were brought individually into a room with a low ceiling, and shown a collection of materials consisting of a large table, two poles about $7^1/_2$ feet long, two poles a little over 3 feet long, a large table clamp and two small clamps, electric bell wire, pieces of lead tubing, and several pieces of chalk. As we shall see,

* Maier draws a distinction between direction and set (241, p. 359), but for our purposes it is one which need not be made.

not all of these materials were used in solving the problem, but the extra material provided cues for false leads. *S* was instructed as follows:

> Your problem is to construct two pendulums, one of which will swing over this point (cross indicated on the floor) and one which will swing over this other point (other cross indicated). These pendulums should be so constructed that they will have a piece of chalk fastened to them which will make a mark (which can be seen) on the points on the floor just indicated (236, p. 118).

The problem of constructing this pendulum is best solved, according to Maier, by clamping two poles together so that their combined length is just the height of the room. Then the two-pole assembly is used to wedge another flat pole against the ceiling. From the ends of the board wedged against the ceiling two strings are dropped over the designated spots. Small clamps are attached to the strings and the clamps hold the chalk.

According to Maier the solution has three essential parts: (1) the concept of a plumb line, (2) the concept of making a long pole by clamping two short ones together, and (3) the wedging principle. In some of the groups of *S*s these parts were indirectly demonstrated before solution was attempted. For example, to demonstrate the concept of a plumb line, *E* showed *S* how to make a plumb line out of a piece of string and a clamp, but instead of chalk a pencil was inserted in the clamp. The other two parts could also be demonstrated. We shall call this procedure, the *part demonstration*.

The instructions used to give direction, or set, were as follows:

> I should like to have you appreciate how simple this problem would be if we could just hang the pendulums from a nail in the ceiling. Of course, this is a not a possible solution but I just want you to appreciate how simple the problem would be if that were possible. Now that it is not possible the problem is, as you may find, really quite difficult (236, pp. 119–120).

Five groups were used as follows:

Group I: Given problem only.

Group II: Given problem plus part demonstration, but told that the part demonstration was given merely to make them acquainted with the materials.

Group III: Given problem and part demonstrations and told that if these parts were properly combined they would yield the best solution to the problem.

Group IV: Given problem and the set instructions quoted above.

Group V: Given problem, part demonstrations and told that parts combined would give solution, and also given set instructions.

This problem proved to be difficult. There were 62 *Ss* in the first four groups; *only one* solved the problem after working as long as desired, which in some cases was 3 hours. In Group V, however, 8 of 22 *Ss* discovered the correct solution. It appeared that not only did *Ss* have to be given part demonstrations and be told that putting certain parts together would give a solution, but they also needed direction of the application of these parts. *Ss* who failed to solve the problem no doubt had a set, but they had the wrong set for this problem.

2. *Experiment II.* The second experiment (237) shows how behavior may be channeled in specific directions by the use of non-verbal as well as verbal hints. Two strings were suspended from the ceiling at such a distance apart that *S* could not simultaneously reach both. The problem was that of tying the two strings together. Again, a multiplicity of objects was present, and actually there were many possible solutions. Maier, however, was interested in only one solution, namely, the tying of a weight (pliers) to one string so that it would act like a pendulum. This would allow *S* to hold the non-weighted string and catch the other as it reached the top of an arc in its swing. If *S* achieved another solution *E* instructed him to keep working for others.

Ss were first allowed 10 minutes in which to discover the correct solution and then, if they had been unsuccessful, two hints were given: (1) *E* brushed across one of the strings to make it swing, and (2) he handed *S* the pliers, telling him that with the pliers and nothing else the solution could be obtained.

Without hints 39 per cent obtained the correct solution in the 10 minutes allowed; after the hints a further 38 per cent solved the problem correctly and 23 per cent failed to find the solution. These results indicate that specificity of set aided solution for 38 per cent of the *Ss*.

3. *Experiment III.* Maier noted in his experiments that many *Ss* seemed to get in solution-ruts or sets which were difficult to break. It was as if *S* felt that his method of attack was correct but he hadn't quite obtained the proper combination of objects. As we have indicated before, whenever there is a problem situation in which the correct response must be discovered, *S* is more likely to discover the response, statistically speaking, in proportion as the number of different attacks tried becomes greater. The rigid sets which Maier and others have noted are just the opposite of the variable behavior which seems necessary to find proper solutions. Maier's observations indicated that those who solved his problems most readily were those who showed varied attacks on the problem; they were able to change the direction of their behavior when one direction proved inadequate. Poor solvers invariably persisted in one line of attack despite repeated failure—a sort of bulldozer response. Maier designed an experiment (238) to determine whether it would be possible to counteract the stereotyping of sets.

Two groups of Ss were used. One was given a 20-minute lecture on problem solving and the other group (control) was given no lecture. The lecture covered general aspects of problem solving with special emphasis on the need to change sets, sometimes change them completely when one method of attack proved fruitless. Each *S* was given three different problems to solve: (1) the string-tying problem as used in Experiment II; (2) constructing a hat rack by clamping two boards together, wedging them against the ceiling, and using the clamp bolt as the hook; (3) blowing out a candle at 8 feet by making a tube out of various parts which were available.

The results are shown in Table 27. It will be observed that there is no difference between the two groups on the string problem. Maier attributes this to the fact that the control group as a whole worked longer on this problem because they had greater difficulty in solving it than did the experimental group. As a consequence, they had less time to spend on successive problems. In this experiment there were 178 *Ss* in the experimental group and 206 in the control group. The groups had not been matched. Maier ran subsequent experiments in which the *Ss* were matched and approximately the same results were obtained as those in

Table 27. We note, that while the difference is not great between the two groups, the lecture had some effect. Maier concludes that thinking or reasoning is in part at least a function of overcoming inappropriate sets.

TABLE 27

EFFECT OF INSTRUCTIONS TO "BREAK SETS" ON PROBLEM SOLVING

Problem	*Instructed Group: Per Cent Solving*	*Control Group: Per Cent Solving*
String	50.6	49.0
Hat Rack	28.7	22.3
Candle	68.3	47.3

Data from Maier (238).

Taken as a whole, Maier's experiments show that set is a very significant factor in thinking. A given set shows itself by persistent mode of response. It may be induced by instructions and may be broken by counter instructions. We should also note the continuity of Maier's materials and methods from experiment to experiment. By such systematic studies we shall be able to raise our knowledge of thinking to a level that is commensurate with our knowledge of other fields in psychology.

Instructions in concept formation. We shall report briefly on an experiment by Reed (303). The methods and materials used were the same as reported earlier under the heading *length of material.* The results of the experiment show clearly how a restriction in the amount of necessary discovery and a sharpening of the direction of attention facilitates concept formation. One group of *Ss* was instructed as follows:

I am going to show you a number of cards, one at a time. Each of these will be named by a nonsense syllable, such as *jok*, *bif*, or *hex*. Look carefully at the cards and try to learn as soon as you can the *name of each card.* At first you will not know the names of any of them and I shall have to prompt you. I shall always prompt you when you fail to tell me the name of a card within three seconds after it has been shown. When I have given you the name of a card *repeat it aloud after me* so that I can be sure you understand it. Your work will be finished as soon as you can name each card without any help (303, p. 75).

Compare the above instructions with the following ones which were given a second group of *S*s:

> This is an experiment in learning concepts. A concept, you know, is a word, or idea that stands for any one of a group of things. Thus, the word *chair*, *bird*, or *stone* stands for no particular chair, bird, or stone, but for any one of a group of chairs, birds, or stones. I am going to show you a number of cards, one at a time. Each of these cards will be named by a nonsense syllable, such as *jok*, *bif*, or *hex*, and *each nonsense syllable is a concept*. Look carefully at all the words on the cards and try to learn as soon as you can the *name of each card* and *what it stands for* (303, pp. 75–76). (The remainder of the instructions was the same as those for the other group).

The first group was given no indication of the nature of a concept and was not told that the experiment concerned concept formation. The first group required a mean of 40.9 promptings, the second only 30.7 promptings, to learn to a criterion of one perfect recitation. Although some *S*s in both groups learned the names as a rote paired-associate task, the number of *S*s using such a method was smaller in the concept-instructed group than in the other. The instructions indicating the nature of concepts restricted the range of trial behavior for that group.

Controlling set by controlling past experiences. We have consistently assumed that when *S* is placed in a relatively new situation, his response will be like his past responses to similar situations. We presume that many of Maier's *S*s were acting in a fashion that had been appropriate in other situations, even though that mode of response was inappropriate for the problem presented. We fully recognize how such sets, if inappropriate, may greatly retard thinking where the solutions demand that other methods be used. Until recently, however, we did not have systematic evidence of how a method of attack is carried over from one situation to another. A series of studies by Luchins (225) gives us full confidence in saying that set can be controlled by controlling past experience.

Luchins' technique was to fixate a single method of problem solving by insuring that such a method would be successful, and then giving problems for which the method was inadequate. The problems used were of the nature which often appear in intelligence tests. *S* is given (hypothetically) three jars which will hold a known amount of water. Using only the three jars as measuring

instruments, *S* is to determine how to arrive at a certain amount of water specified by *E*. Thus, if one jar held 5 units, another 13 units, and another 2 units, how would one get exactly 1 unit of water? This could be done by filling the 13 unit container, and from it filling the 5-unit jar twice and the 2-unit jar once, thus leaving 1 unit. Or, it could be done more simply, by filling the 5-unit jar, and then from it filling the 2-unit jar twice.

In his basic experiments Luchins used nine problems. The first was a practice problem, so we will not include it here. The following eight problems are shown below:

	CONTAINERS			
	a	*b*	*c*	*To obtain*
Problem No. 2	21	127	3	100
Problem No. 3	14	163	25	99
Problem No. 4	18	43	10	5
Problem No. 5	9	42	6	21
Problem No. 6	20	59	4	31
Problem No. 7	23	49	3	20
Problem No. 8	15	39	3	18
Problem No. 9	28	76	3	25

To obtain 100 units of water as requested in Problem No. 2, *S* would fill the 127-unit container, and from it fill the 21-unit container once and the 3-unit container twice, thus leaving 100 units in the large container. If these operations are put in algebraic form the equation reads $b - a - 2c = 100$. Now, if you will check problems 3, 4, 5, and 6 you will find that they may be solved in exactly the same way and in no other way. Problems 7 and 8 may also be solved by this equation, but, also, they may be solved in a much more efficient way by not using the b-jar at all. Problem No. 9 can be solved only by a short method (28-3), and is insoluble by the long method.

In a typical Luchins experiment two groups of *Ss* were used. The *Ss* in both groups were given the same problem (No. 1), and then the control group was dismissed. The experimental group remained and solved, in order, Problems 2 through 6. Then both groups together were given Problems 7, 8, and 9. In working Problems 2 through 6, the time allowed was such that most *Ss* were successful in finding the correct solution. The question *E* asked was what influence will the method of solution developed

on these five problems have on the solution of the succeeding problems?

Luchins has given these problems to literally hundreds of *Ss* of varying ages and educational status. The effect of the initial five problems was to produce a carry-over of the method of solution to Problems 7 and 8. The per cent of *Ss* in experimental groups using the old method of solution ranged from 70 to 100 per cent, whereas in the control groups, the per cent using the method was consistently near zero. The effect was different with different age groups, but there can be no doubt about the powerful effect of the transfer of the solution method. In this case, of course, such a transfer was inefficient, since Problems 7 and 8 could be solved in an easier manner.

The results on Problem 9 also show the effects of the set developed in working the first five problems. *A great many* Ss *in the experimental groups were unable to solve Problem 9 in the time allowed* although the solution was *much simpler* than solutions which had been used on previous problems.

Luchins systematically varied several different conditions, some of which we should mention. If *E* increased the number of set problems (beyond the five actually used) the set effect on successive problems was greater. This finding clearly points to the fact that the set was being learned just as specific responses are learned. If *E* asked *Ss* to write "Don't Be Blind" on their papers immediately after the sixth problem, the set effect was reduced somewhat on the following problems. Thus, as Maier's results have shown, the set effect can be reduced by appropriate instruction, in this instance by just giving *Ss* a jolt. This is also demonstrated by procedures in which the first five problems were alternated with other problems not solvable by the same formula. In such instances the set effect was much reduced.* Giving speed instructions increased the effect of the set on subsequent problems; allowing a time interval to elapse between the first five problems and the test problems decreased the set effects. Here we have a phenomenon which may be reduced or increased as a function of certain conditions, but under standard conditions such as were initially described, its effect is very potent. Luchins' results fully warrant his conclusion that the initial problems (2 through

* Harlow (140) has reported a very similar finding with monkeys as *Ss*.

6) create "a mechanized state of mind, a blind attitude toward problems; one does not look at the problem on its own merits but is led by a mechanical application of a used method" (225, p. 15).

We have covered enough experiments to demonstrate the methods by which set may be induced in thinking problems, and we are clearly aware of its importance as a determinant of the rate at which thinking will take place. There are many other types of problems on which influence of set needs to be determined, but tentatively we may identify set (as produced by instructions or past experience) as being especially important in situations in which the number of apparent solutions to a problem is great.

The Effect of Unanalyzed Biases

In an earlier chapter certain experiments were described which showed the influence of unanalyzed biases on the discriminal processes. Since the thinking situation is often one which requires discrimination among various stimuli, it might be expected that our responses will be determined in part by the unanalyzed biases which influence our perception. These factors may be conceived of as sets which have their origin in learning processes in a manner which Luchins so clearly demonstrates. However, because *E* does not have control of the conditions under which the biased responses arise, one may continue to speak of them as unanalyzed biases. Their importance in thinking is the same as the set factors which were analyzed and traced to sources; they influence the rate of thinking, since they determine in part the responses which *S* makes to a thinking problem. Three illustrations follow.

Atmosphere effect. Woodworth and his students at Columbia University have discovered an "atmosphere effect" which influences responses to syllogisms. The effect is shown by analysis of the acceptance of various invalid conclusions. If a syllogism is stated in affirmative form, *S* is more likely to accept an affirmative than negative conclusion, even though both conclusions are invalid. Likewise, if a syllogism is stated in negative form, the negative conclusion is more likely to be accepted. Take, for example, the following premises:

If all x's are y's
And if all z's are y's

The "atmosphere" hypothesis predicts that *S*, if he accepts an invalid conclusion, is more likely to accept a conclusion which is affirmative in tone. Two conclusions which would be affirmative in tone would be:

Then all x's are z's
Then some x's are z's

S would be very unlikely to accept the following two conclusions which are negative in nature:

Then no x's are z's
Then some x's are not z's

Of course, all four conclusions are invalid from the premises. Yet, when *S* does accept an invalid conclusion as true, as Sells' (345) work shows he often does, the conclusion he accepts has the same "atmosphere" or "impression" as the premises. If the premises were negative, such as:

If no x's are y's
And if no z's are y's

S would be much more likely to accept an invalid conclusion stated in a negative fashion than he would one stated in affirmative fashion.

It is quite clear from Sells' work that a bias operates to determine in part one's choice of conclusions.

Attitudes. Some scattered evidence suggests that logical thinking may be distorted by personal attitudes.* However, a study by Morton (271) is the first which clearly demonstrates that emotionally toned attitudes may markedly bias our responses to syllogisms. There are two important parts to the study. One part concerns the acceptance of conclusions from premises stated in the symbolic *x* and *y* form. A second part concerns the acceptance of conclusions when the premises and the conclusions were stated in terms of current events; namely, events of World War II (the study was done in 1942). These syllogisms in the second part were couched in terms of rationing, the airplane versus the battleship, spies, enemy leaders and soldiers, and so forth, all of which

* The relationship between emotion and problem solving was taken up in Chapter VII.

were very much in the news at the time and about which most people held strong opinions.

The test, then, consisted of syllogisms in formal symbolic language and syllogisms in emotionally-toned items, but both sets of syllogisms were identical as far as logic was concerned. There were 15 syllogisms, each stated in the two ways. *S* was asked to mark one of five possible conclusions which he thought most logically followed from the premises. Then *E* compared the responses for each syllogism stated in formal terms with the responses given to the same syllogism when stated in terms of current topics.

The responses for the 98 *S*s (college students) were different for the two sets, and these differences were highly significant in 14 of the 15 syllogisms. The most important finding was that the logically wrong conclusion accepted most frequently was one which represented the prevailing opinion by newspapers, radio, and other forms of public commentary. Thus, the source of this bias seems fairly well identified. In short, a conclusion which reflected public opinion was likely to be accepted, although from the premises such a conclusion was quite illogical. (This does not necessarily mean that the accepted conclusion was not true in and of itself, e.g., "The Japanese advances in the Aleutians can be held," but it does mean that such a conclusion was utterly illogical in the context of the given premises.)

Morton also showed that when the emotionally-toned conclusion in line with current opinion was opposed to the atmosphere effect (as discussed above), the emotionally-toned statement was more likely to be accepted than the one predicted by the atmosphere effect.

Size and position biases. A study by Morgan (270) illustrates in a very convincing manner how irrelevant factors may be given causal status in problem solutions. Morgan presented his *S*s with a situation which may be schematized as follows:

A	B	C
A	C	B
A	D	C

The letters represent levers and *S*s were told that when all levers in a row were pressed a door would open. Then *S* was requested

to rank-order the four stimuli (A, B, C, D) as to their importance in causing the door to open. Now you will note that in the above illustration both A and C occur in each row so by common logic they should be equally important in producing door-opening. Morgan reasoned that if students applied logic only they would be unable to rank either A or C ahead of the other. The question was, then, what irrelevant factors would determine a choice between A and C?

In the actual experiment four geometrical forms (a circle, a diamond, a square, and a triangle) were used instead of letters. These forms appeared in various sizes and at various positions in the 12 problems used. Some forms had lines through them. Each problem was exactly the same; *S* was to rank-order the four stimuli as to their importance as causal factors in opening the doors.

Out of 246 college students who served as *Ss*, only 6 absolutely refused to rank-order the stimuli, rightly insisting that the two factors were equally important and one could not be ranked ahead of the other. Responses of the other 240 *Ss* were examined to determine which factors were chosen as being important. The results show that two factors, size and position, stood out above all others as alleged causal conditions. Practically all *Ss* ranked the figures appearing in all rows either first or second, but a figure was much more likely to be ranked first if it were larger than the other or if it were in the dominant position (such as the A's in the illustration).

The mean ranks for all characteristics are shown in Table 28. It can be seen that little bias attached to a given geometrical form as such and whether or not the figures were plain was of small consequence. The important point is that *Ss* preponderantly showed a bias toward certain features (size and position) even though those features are irrelevant causal factors in this problem.

We have given three experimental illustrations of unanalyzed biases which may influence our responses in thinking. These will give an inkling of the possible sources of the irrelevant or inadequate behavior which is often encountered in problem solving, our own and others. Though biases are no doubt due to sets invoked by past experiences, tracing them to their sources is a difficult task. It is certain that the biases thus far experimentally

isolated are far outnumbered by those which have yet to be found.

TABLE 28

EFFECT OF NON-RATIONAL FACTORS ON THINKING

	Mean Ranks			
	Square	Circle	Triangle	Diamond
Dominant Position	1.12	1.16	1.14	1.21
Non-Dominant Position	1.83	1.89	1.88	1.87
Large Size	1.29	1.32	1.30	1.24
Small Size	1.73	1.71	1.77	1.73
No Lines Across Face	1.25	1.66	1.40	1.71
Lines Across Face	1.36	1.66	1.30	1.77

Data from Morgan (270).

Memory Factor

The work on concept formation has suggested that concepts are more difficult to attain if *S* has to draw on memory to supply the characteristic defining the concept. Thus, if the correct concept were animals; and cows, horses, pigs, and elephants were pictured together, the common factors among the pictures are evident. Let us suppose, however, that *S* is to form a concept of *transportation*. *S* is told that pictures of cows, sheep, buildings, and desks do not illustrate the concept (*E* reports "wrong"), whereas pictures of horses, cars, elevators, and boats do illustrate the concept ("right"). Nothing concerning transportation is given in the immediate perception of the pictures; *S* has to import from memory this functional characteristic which holds for all of one group of objects and not for the other.

Such evidence as is available points toward the principle that the greater the dependence of thinking on memory, the slower is the rate of thinking. One pertinent experiment by Maier (241) may be used to demonstrate the probable existence of such a relationship.

Three groups of 25 *S*s each were used. In the description to follow you will again note the continuity of the materials with those used in the previously discussed experiments.

Group I. *E* and *S*, working together, constructed uprights by clamping boards together. Two sets of two uprights were made, each set being used to wedge another board against the ceiling. From each of the wedged boards two strings were suspended, far enough apart so that the tying-strings problem was presented to *S*. If *S* did not succeed with the problem in 10 minutes, *E* showed him the correct solution. When this problem was completed, *S* was taken to the other end of the room and given two poles, 6 feet and 7 feet in length and instructed to build a hat rack strong enough to hold a heavy coat and hat. The solution to this problem, we have seen, is provided by clamping two poles together and wedging them vertically between floor and ceiling so that the arm of the clamp serves as the coat rack.

Group II. The same experiences were given *S*s in Group II as were given those in Group I. However, the structure used to suspend the strings was taken down and the parts removed from *S*'s vision before he was presented the hat rack problem.

Group III. This group of *S*s worked only on the hat rack problem—they were given no preliminary experience in building uprights by clamping boards together.

TABLE 29

INFLUENCE OF THE MEMORY FACTOR ON THINKING

Group	*N*	*Number Failing*	*Per Cent Failing*
I	25	7	28.0
II	25	13	52.0
III	25	19	76.0

Data from Maier (241).

The results of the attempts to construct the hat rack are given in Table 29. The data show clearly that Group I was superior to Group II, and Group II superior to Group III. This experiment demonstrates: (1) that past experience in solving a problem with

a given material will facilitate the solution of other problems using the same or similar principles and materials, and (2) that the facilitation provided by the past experience is greater if the stimuli associated with the past experience are given in the immediate perceptual field so that *S* does not have to rely on his memory.

It must be understood, of course, that experiences are not remembered without stimuli. They are recalled on the basis of current stimulus patterns, although we are often hard-pressed to point out the stimulus which evoked the recall of a past experience. However, if the experience recalled is inappropriate to the current problem situation, solution of the problem will be delayed. Maier's experiment shows that appropriate experiences are recalled most quickly when the stimuli directly associated with the experiences are present.

A somewhat related principle may be derived from Köhler's well-known work with chimpanzees (204). If several objects must be brought into functional relationship before a problem can be solved, this functional relationship is more likely to be perceived when the objects are spatially contiguous. Thus, if a chimpanzee had to stack two boxes in order to reach a banana, the chances of his being able to do this were increased when the two boxes were placed close together under the banana. If the first box were at one end of the case and the second at the other end, the correct stacking response was less likely to occur.

We may extend this contiguity principle of spatial objects to memory relationships. If the solution of a problem is dependent upon *S*'s observing the functional relationships between several past experiences, any condition which tends not only to evoke memories of the pertinent experiences, but also to produce temporal contiguity of the memories, should facilitate problem solving. Unfortunately, such a statement has the status of an hypothesis since there are no data available to provide its verification. Note that the principle does not state that sheer recall of appropriate experiences automatically means that a problem is solved. Maier's work shows that this is not true. However, temporal contiguity in the recall of appropriate experiences should increase the probability that *S* will perceive the functional relationships among the stimuli involved in the problem.

MISCELLANEOUS PROBLEMS

Animal Studies

We shall describe briefly certain animal studies which may be considered to represent more complex forms of learning in animals. If there is true "thinking" in animals, these studies probably represent it.

Delayed reaction experiments. In delayed reaction experiments *S* is shown two or more containers and observes *E* as he places food under or in one of the containers. *S* is not allowed to approach the container immediately and a barrier is placed between the animal and the container for the delay period. The basic problem concerns the length of time the animal can be delayed between his perception of *E*'s putting the food under the container and his making the correct response. Studies have been directed at an evaluation of the important variables which influence the length of successful delay.

Combining past experiences. Maier has attempted to determine whether rats can combine two separate experiences into a single experience. The typical apparatus involves three platforms placed so that they form the corners of a triangle. Elevated alleys run from each platform to make a junction in the middle of the triangle. Thus, any platform may be reached from any other via the elevated alleys. At the start of the experiment the rats are given time to explore the alleys and three platforms. On a test trial, the rat is given a bit of food on one platform, then carried to another. Will the rat be able to combine past experiences so that it will run down to the junction and up the alley to the platform where food had been received, avoiding the other alley? It will (235).

Discrimination learning. There have been many experiments on discrimination learning, using various types of apparatus and animals. We are already acquainted with certain of these techniques, such as the jumping technique and the T-maze. A recent publication by Harlow (141) gives a comprehensive survey of the various methods which have been used and some of the findings concerning stimuli (form, size, color, etc.) which are most easily discriminated.

There are certain theoretical controversies regarding the mechanisms by which discrimination learning is achieved. Since these

controversies are reflected in the theories of learning to be discussed in the next chapter, we shall not consider them here.

Discrimination learning usually involves the training of a positive response to one stimulus and a negative response to another. The method has sometimes been called the *method of contrasts*. *Transposition experiments* are studies in which animals are initially trained to make a discrimination between two stimuli. *Ss* are then tested on two stimuli having magnitudes of the same ratio as the training stimuli, but differing in absolute value.

Descriptive Analysis

We have not discussed the many introspective reports on thinking which perhaps show something of the thought processes *S* goes through in reaching a solution to a problem. Previously, we have mentioned that these studies have not provided us with appreciable systematic knowledge regarding causal factors in thinking, i.e., they have added little to our ability to predict.

One of the most recent extensive descriptive analyses of thinking, from a subjective point of view, is found in a book by Wertheimer (404). In this account he differentiates between rote application of past experience in a problem situation, on the one hand, and on the other, the application of past experiences with understanding of the relationships among those past experiences. His major aim is to show how "true" thinking results in a unique response—a unique perception of inter-relationships among experiences. For a modern descriptive analysis of complex thinking, Wertheimer's book should be read.

Is Thinking Learning?

We have arbitrarily called thinking a form of learning. The modification of behavior which occurs in thinking meets all the definitional requirements of learning. Yet, we would be misinterpreting the status of thinking in American psychology if we left the problem at that point. Two facts should be pointed out: (1) very few psychologists are concerned about whether thinking is or is not learning, and (2) those few that are, vigorously assert that thinking—"true" thinking—is more than learning. Maier is the chief proponent of the latter point of view.

Maier (240) does not conceive of many of the tasks we have

covered in this chapter as requiring true thinking. Concept formation, for example, is explained by Maier through a quite different set of principles than is thinking, as he has studied it among rats and college students. The term *thinking*, according to Maier, should be restricted to situations in which there is spontaneous integration of past experiences; this sets it off from all other forms of behavior modification. This is not speculation on Maier's part, since he adduces considerable evidence for his point of view.

It is not our plan to attempt to reach a decision on this matter. We need only know that there are different points of view on this problem; at least, there are potentially different points of view, since many psychologists who have not committed themselves on the matter in print would probably disagree with Maier. The question which must be ultimately answered is whether or not the basic principles which explain learning (conditioning, multiple-response) may be used also to explain thinking.

FUTURE TRENDS

We have seen that our knowledge of the relationships between manipulable variables and rate of thinking is meager and tentative. Rather than give illustrative experiments for certain variables, we have had to give the only experiment available. We have said that the amount of current research being done on thinking is small as compared with the amount being done on simpler forms of learning. Now we may ask, "Where do we go from here?"

It seems what is needed is vigorous and widely-joined theoretical controversies centered around how people think. The current temper of scientific psychology is such that theoretical formulations point directly to potential experiments; if we had the theories they would certainly spawn research. It is to be noted that the few psychologists who have shown sustained research interest in thinking (e.g., Heidbreder; Maier) have had theories which they were testing. We need more psychologists with theoretical insights in order to bring the status of research in the area to a respectable level.

One of the initial needs for future research is a series of standardized tasks representing several levels of complexity. If *E*s did their work with such tasks, inter-laboratory communication would

be greatly facilitated. The *systematic* exploration of the influence of variables—all variables—for a variety of tasks, remains as the major empirical program.

SUMMARY

1. Thinking has been identified as representing the most complex form of learning, involving a maximum of discovery of relationships among stimuli. Rational learning problems were taken as a representative transition task between multiple-response learning and thinking.

2. In general the research on thinking is retarded as compared with the research on simpler forms of learning. Several reasons for this state of affairs were discussed.

3. The following variables were shown to influence the rate of thinking:

a. Massed vs. distributed training
b. Type of material
  1. Length of task
  2. Nature of concept
  3. Intra-task similarity
c. Set
d. Memory factors

4. Related work with animal studies was briefly outlined, and differing points of view concerning the relationship between learning and thinking were mentioned.

CHAPTER XIV

Learning IV — Theoretical Problems

INTRODUCTION

The writer once heard a story concerning an honored society of aged gentlemen which met once a week to exchange views and argue on various issues. It so happened that one of the members died and it was necessary to replace him. After assessing the various candidates, they finally chose a relatively young man who seemed to be wise in the ways of the world, at least verbally, and who they thought would fit nicely into their group. The first meeting which the young man attended was dedicated to the discussion of the probable number of teeth in a horse's mouth. For an hour or so each of the elder gentlemen had many wise words to say concerning the number of teeth which he believed a horse had. In deference to his elders, the new member of the society spent most of the session listening. Finally, however, in a show of temerity, he baldly suggested that the meeting retire to the stable where it would be possible to count the number of teeth which a horse actually possessed. The elders, aghast at such an approach to the problem, immediately dismissed him from their circle and asked that he never return. The appeal to the empirical was not a part of their mode of thought.

Contemporary learning theories are oriented by facts. Only after a horse has been observed is there a theory; the theory follows and is based on the observation of number of teeth. The sequence of events in formulating scientific theory is: facts → theory concerning facts → more facts to check on adequacy of theory → revision or extension of theory → and so on. Unlike the gentlemen in the story, the modern learning theorist neither starts nor stops with theory; fact finding and theory construction are

alternate, closely integrated steps. Nearly all current experiments on learning are designed to test directly or indirectly certain aspects of learning theory. In this chapter we propose to sample differences in points of view concerning the basic nature of learning, and to sample experiments which are presumed to give critical tests of these theoretical points of view.

PRIMARY AND SECONDARY VARIABLES

In three previous chapters we examined factors which were effective in determining the rate at which learning took place, and we called these factors the *effective stimulus variables* or *dimensions*. These variables were discussed for their rôles in simple learning situations (conditioning) and for complex learning situations (thinking). Thus, meaningfulness, complexity of conditioned stimulus, intra-task similarity, distribution of practice, and so forth, were all shown to influence the rate of learning.

We have never spoken of conditions which are *necessary* or *essential* for learning. Thus far in our presentation we have tacitly assumed that learning is a fundamental, almost inevitable capacity of the organism. We now ask: "Are not certain conditions essential before any learning will take place?" "Are there not situations in which no learning will take place *in spite* of optimal distribution of practice, or optimal conditions of meaningfulness, or optimal conditions of intra-task similarity?" It is the purpose of this chapter to evaluate various hypotheses which are incorporated into learning theories and which are presumed to answer such questions as those above.

A classification scheme. Let us call the experimental variables which we discussed in the preceding three chapters, *secondary variables*, and the empirical relationships between these stimulus variables and rate of learning, *secondary laws or principles*. We shall then call the variables to be discussed in this chapter *primary variables* which lead to *primary laws of learning*. (Two variables discussed in the previous chapters which would not be secondary are the temporal relations in conditioning and the ratio-of-reinforcement variable. As will be seen presently, these two represent variations in aspects of primary variables.)

We shall see that a distinction between primary and secondary

variables is not completely justified, but it may aid us somewhat in handling the rather difficult problems we face in completing the study of learning. There are four variables which rank as primary: (1) motivation, (2) consequences, (3) frequency, and (4) contiguity. The central problems of current learning theory center around the mode of action of these four variables.

1. *Motivation.* Common observation suggests that motivation is a necessary condition for learning. Even to attend to stimuli the organism must be motivated. No learning theorist insists that *appreciable* learning does take place in non-motivated organisms, although there are differences of opinion as to the exact rôle which motivation is to be assigned in the learning process; some theorists say learning *could* take place in non-motivated organisms. The operation of motivation is closely bound up with the operation of incentives in most learning theories.

2. *Consequences.* As a result of our discussion in Chapter VI, we are already acquainted with some of the problems related to incentive variation. The incentives are the rewards and punishments; they are the *consequences* (or anticipated consequences), of a response or act. We should conceptualize the consequences of an act in terms of a continuum, with one half of the continuum indicating rewarding consequences and the other half punishing consequences. There would then be a neutral point, or zone, in the middle.

Neutral

Rewarding Consequences : *Punishing Consequences*

From the neutral point out toward the ends the amount of punishing or rewarding incentive increases. At the empirical level the influence of "amount of consequence" on learning is a straightforward experimental problem. *E* varies the amount of reward (or punishment) and measures the rate of learning which results for each variation. It should be noted, however, that we cannot define the nature of consequences in terms of their influence on rate of learning. If a consequence hastens learning we cannot therefore define it as a reward. We have to define the dimension independent of the rate of learning.

Definition of the dimension of consequences has been achieved on the verbal level by Thorndike (380). A consequence to which

the organism shows approach responses would be a *reward;* a consequence to which the organism shows avoidance responses would be a *punishment.* If the organism neither avoids nor approaches a given incentive, it would be classed as a *neutral consequence* (assuming no ambivalence or conflict). Strength of approach and avoidance responses would indicate the amount of reward and punishment respectively.

We will recognize that most approach and avoidance responses are in themselves learned so that a definition of consequences is not completely independent of observations of learning behavior. However, the incentives which are typically used evoke fairly stable responses the learning of which has long since taken place.

In actual practice the pre-testing of the nature of consequences is seldom carried out, although it could be; a rat would approach food, and avoid shock. Given these independent observations for a given motivational state, we could then proceed to discover what influence such incentives had on learning. The scientist generalizes to members of a species from a sample and since it is common observation that rats approach food and avoid shock, he does not feel obliged to test each rat for its reactions before an experiment. If incentives do facilitate or retard learning, it is a part of the theorist's problem to make assumptions as to how this facilitation or retardation takes place.

3. *Frequency.* The frequency variable might seem so obvious that its importance could easily be overlooked. We already know that at least a frequency of two repetitions is necessary before learning can be measured and we also know that it is customary to plot learning curves by placing frequency (number of trials) along the base line. However, frequency becomes a theoretical problem when we pause to inquire *what is frequent.* What *is* frequent that is critical for learning? The response? The stimulus and response combined? The response and reward? Some theories have fundamental disagreements over the answers to these questions.

4. *Contiguity.* We have already examined some of the empirical relationships between rate of learning and contiguity of certain aspects of the learning situation. There are differences of opinion concerning the relative importance of certain "contiguities": Is the temporal interval between stimulus and response the critical

one, or is it the interval between response and reward? We call this variable primary because empirical data show that if the intervals between stimulus and response, and response and consequence are too long, no appreciable learning will take place.

Let us schematize the relations between primary and secondary variables by the structure in Fig. 75. This structure shows each primary variable underlying all secondary variables. By this is implied first of all that the primary variables must be present in certain minimal amounts before learning will take place. Increas-

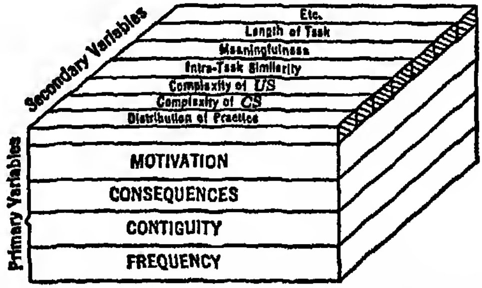


FIG. 75. A structure to schematize the dependence of secondary variables on primary variables.

ing the amount of the primary variables beyond this minimum also affects the rate at which learning takes place. The exception is *frequency*, since frequency is necessary only as a vehicle to allow learning and to measure rate of learning. In a sense, frequency underlies all variables since a minimum frequency (repetition) of one sequence of specified events must take place before learning occurs, and two frequencies before learning can be measured. The secondary variables are entirely dependent on the primary variables—if we remove the primary variables, manipulation of the secondary variables would have no influence since learning cannot take place without the primary variables. However, given at least a minimum amount of all of the primary variables, the rate of learning will be influenced by manipulation of the secondary variables.

It must be clear that not all theorists will agree with the classification of primary and secondary variables. Some of the variables which here appear as secondary would be placed by some authors

in the primary classification, whereas certain of the primary variables might be placed in the secondary classification. For example, some theories which strongly stress the perceptual aspects of learning would probably include intra-task similarity as a primary variable, since it influences the ease of perceiving various aspects of the situation. Other theorists say that incentives (consequences) are not necessary for learning to occur. In spite of disagreements over the classification of primary and secondary variables, current psychological writing shows that the major theoretical problems stem from different preconceptions concerning the mode of action of the primary variables as here listed.

Let us turn briefly to some general aspects of theory construction and follow with a discussion of the rôles assigned the primary variables in different theories.

ILLUSTRATIVE THEORIES OF LEARNING

Theories are attempts to give explanations of observed phenomena on the basis of given empirical relationships. For example, distributed practice usually facilitates learning—this is an empirical relationship. The theorist attempts an explanation by suggesting processes (and their characteristics) which might produce the ascertained empirical relationship.

Theories and scientific development. Theory building usually leads to an intermediate stage in the development of a science. As pointed out in the first chapter, theories carry experimentation in a given area from the I-wonder-*what*-would-happen stage to the I'll-bet-*this*-would-happen stage. The scientist, having established some empirical relationships in the first stages, becomes curious about the underlying processes which could produce such a relationship. Consequently, he begins to propose possible explanations or hypothesize. And, as we also pointed out in the first chapter, if the theory is to be maximally useful, it must suggest relationships which have not been observed as yet but which are open to investigation. Further experimentation thus takes the form of seeking for empirical relationships which are suggested by the theory, and we have the I'll-bet-this-would-happen stage. Our confidence in the basic theory is strengthened if predictions are verified, and weakened if they are not verified.

In earlier chapters research areas were mentioned in which investigation has virtually ceased for want of a theory. At these points no one has formulated a theory which would push experimentation from the I-wonder stage to the I'll-bet stage. Whole-part learning is a small area which seems to be moribund for want of theory. Thinking is a larger area which as a whole seems to be stalemated for lack of a comprehensive theory.

Levels of generality. It is agreed that theories may be of several levels of generality. A theory may be formulated to account for only a very restricted segment of behavior, or it may be elaborate enough to account for a wide range of behavior. Thus, a very restricted theory would be one which attempted to explain the bowed serial-position curve. A less restricted theory would be one to interpret the bowed serial-position curve, and remote associations, and the superiority of distributed over massed practice. A still more general theory might attempt to account for all multiple-response learning phenomena. Theories which have been devised to explain the smaller areas of behavior are sometimes called *miniature systems*.

The nature of learning theory. Ideally, a theory of learning would account for all known empirical relationships. It is perhaps unnecessary to state that we have no such broad theories. Most theorists have restricted themselves to building miniature systems. Yet, a few have attempted to account for a fairly wide range of behavior, and it is their work which we propose to sample in the discussion to follow.

The theories to be discussed have evolved largely from animal experimentation but are intended ultimately (with elaboration) to apply to all learning. Animal experimentation has provided a convenient means for isolating certain variables, especially those concerned with motivation. We shall not deny, however, that there are difficulties encountered when one tries to apply to human learning the theories derived from white rat learning. However, the problems which the theories raise are by no means problems of animal behavior only. Many human experiments have taken their cues from animal experiments, and we shall find many incidents in our own experiences which seem to illustrate the basic assumptions made by the theorists. Although these personal illustrations do not validate the assumptions, they may

alleviate the feeling that the theories are exclusively theories of animal learning.

We shall illustrate learning theories by considering (1) a goal-reinforcement theory, (2) a perceptual-learning theory, and (3) a contiguous-conditioning theory. It would be presumptuous to believe that it is possible to outline these theories fully in a few pages, when originally it took several volumes to present them. We shall, however, be able to state the basic assumptions of each theory in simplified form and point out differences in them. A recent book by Hilgard (151) reviews in some detail a dozen different learning theories which influence modern psychology, so it is clear that the three theories we consider here are only samples.

Goal-Reinforcement Theory*

This theory is given its most recent and extensive treatment by Hull (165). The theoretical system as it stands today is based largely on conditioning data—unitary response learning. Future volumes are promised to provide a theoretical framework for multiple response acquisition, especially maze learning. The theory is thoroughly grounded in experimental data, and is stimulating a great amount of research.

The key assumption of the theory is a modern statement of the law of effect which we can approximate for the present by saying that the consequences (rewards or punishments) of a stimulus-response sequence are critical in determining whether or not learning takes place. Hull has borrowed most directly from Thorndike (378, 380) for his ideas on the rôle of effect. There are 16 basic postulates in Hull's system and many corollaries, of which we shall include only the more critical ones. The theory will be outlined in terms of the four primary variables previously discussed, although we shall treat motivation and consequences together. Experiments designed to test propositions deduced from the theory will come under our attention later.

* To summarize these theories the writer has used original sources (cited in text); interpretative material by Hilgard (151), Spence (362), White (405); the discussions in the 1942 *Yearbook of the National Society for the Study of Education* (147), and many discussions accompanying experimental reports.

Motivation and consequences. There are three points to be considered.

1. Motivation, in the energizing sense, must be present before learning will take place.

2. Whenever a stimulus and a response occur together, and this in turn is followed closely by a reinforcing state of affairs, the functional connection between the stimulus and response is strengthened (law of effect). By this it is meant there is an increase in the probability that upon future presentation of the stimulus the response will occur.

3. A reinforcing state of affairs is brought about by two conditions:

a. By the reduction of a motive (a pellet of food reduces strength of hunger motive; cessation of shock reduces escape motive). This is primary reinforcement.

b. By stimuli which have been associated with primary reinforcement. This, we know, is secondary or symbolic reinforcement. In humans, Thorndike's confirming or "O.K." reaction may be thought of as secondary reinforcement. Thus *S*, saying the response *GUZ* to the stimulus *YOR*, and then seeing that *GUZ* is correct or "O.K.," has his response confirmed (reinforced).

Frequency. Since each occurrence of the stimulus-response sequence when followed by reinforcement increases the probability that the stimulus will in the future evoke the response, level of learning is closely related to frequency. The important frequency is that of stimulus-response-reinforcement sequence, not merely frequency of stimulus and response pairing. Each occurrence of this sequence builds up cumulatively a certain amount of associative strength between stimulus and response.

Temporal contiguity. Two aspects of this variable are important for goal-reinforcement theories: (1) the temporal relations between stimulus and response in Procedure I (Pavlovian situation), and (2) the temporal relations between response and reward in Procedure II (Skinner-box). The data in Chapter XI have shown that in Procedure I the optimum interval between stimulus and response is somewhat less than 1 second, and for Procedure II simultaneity of response and receipt of reward is optimal for learning.

Perceptual-Learning Theory

The perceptual-learning theory has been given its most elaborate statement by Tolman (386). Whereas Hull's theory is based largely on conditioning situations, Tolman's theory stems primarily from simple and complex maze learning with rats. Because the theory is grounded in experimental work, because certain predictions which stem from the theory are testable, and because it stands in opposition to goal-reinforcement theories, it too is currently provoking considerable experimental work.

This theory differs basically from a goal-reinforcement theory in that it assumes rewards *do not strengthen functional connections*. Rewards only "call out" behavior which depends on the organism's perception of the situation. Acquiring this perception, not acquiring responses, is learning. Rewards do not strengthen directly.

Learning, according to the perceptual learning theory, takes place when an organism perceives a stimulus and its significance. A rat in a simple T-maze will, after a few trials, perceive that the stimuli at the choice point indicate that "down that alley" there is a stimulus (food) which has significance for him. Food, as opposed to no food, does not increase the response tendency by reducing motivation; instead, it aids in sharpening the perception of the relationship between the stimuli of the choice point and the stimuli of the goal box. Other factors which aid perception will also help in the formation of this relationship.

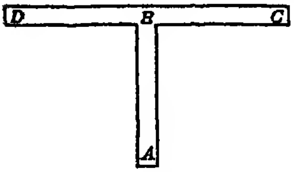
Learning thus becomes equivalent to "seeing" that such and such a stimulus followed by such and such behavior (running) will lead to the goal box. Learning is the acquisition of certain bits of information about the relationship of various stimuli in the environment, so that one stimulus arouses the *expectation* of others. Learning is thus the establishment of certain *sets* or *dispositions* to act in certain ways when the significant stimuli are present. Factors which bring about learning are factors which increase the clarity of the perceptual relationship among stimuli in the environment.

If it appears puzzling to us that rats have all these capabilities, let it be said in Tolman's defense that most of the terms are adequately defined in terms of experimental operations. But now, let

us be more systematic and precise with this theory by discussing it in terms of the four primary variables.

In outlining this theory, let us consider a simple T-maze in which the starting point is *A*, the junction is *B*, and the ends of the two arms *C* and *D*. As we advance point by point through the theory, we shall refer to various events which are assumed to occur in the T-maze.*

Motivation and consequences. 1. Suppose we could put a completely non-motivated animal in the maze. If this could be done, the rat would show very little activity. According to this theory, motivation is necessary for learning, but only because it arouses the organism to explore the environment and attend to stimuli.



2. Next, suppose we put food in the box at *C* and use a hungry rat as *S*. In the succeeding trials the rat quickly learns to go to *C* from *A*. The goal-reinforcement theorist says that this increase in performance is due to the fact that food serves as a primary reinforcement to strengthen the stimulus-response sequence which preceded it. The perceptual-learning theorist, on the other hand, maintains that what actually happens is that the rat experiences stimuli at *A* followed by stimuli at *B*, and thus builds up a bit of information that *A*, followed by appropriate locomotion, leads to *B*. Furthermore, *B* becomes experienced contiguously with *C*; hence, another bit of knowledge is established that *B*, followed by running in "that direction" leads to *C*. Since we presume that the animal occasionally went from *B* to *D*, the information that *B* followed by running in "that direction" leads to *D* is also built up.

Now, this learning of stimulus relationships could take place without food in the box at *C*. If we let a rat explore the maze for a period of time before placing the food at *C*, we should expect, by this theory, to find its performance following introduction of food considerably superior to that of a rat which had not had the exploratory trials. In short, perception of the relationships

* No implication is intended that other theories, e.g., goal-reinforcement, cannot account for the simple behavior phenomena to be discussed. The terms *prediction*, *account for*, etc., are used for expository purposes.

among stimuli *is* learning, and these percepts can be acquired, according to this theory, in the absence of food. In the maze the pertinent perceptual relationship concerns the *spatial relations* of various stimuli both within and outside the maze. Placement of food at *C* following exploratory periods allows learning which took place during exploration to show itself in the performance.

Why does the rat learn the path to food? Why doesn't it take the path to *D* as often as it does to *C*? The answer is that food serves to distinguish the stimuli at *C* from those at *D*. With food at *C*, the rat learns that stimulus *B* followed by running in that direction leads not only to *C* but also to *C* and food. Thus, this bit of information is distinguishable from the *B* to *D* relationship, since *D* does not include food. Food helps to sharpen differences among the two sets of stimuli, and if the animal is hungry, it will go to *C*.

Frequency. To account for learning curves, frequency becomes a critical variable for a perceptual-learning theory. However, what this theory holds important is that stimuli frequently appear in close relationship, and not, as the goal-reinforcement theorists say, that stimulus-response-reward is frequent.

Contiguity. This theory stresses contiguity *between stimuli.* The more temporally contiguous is the perception of two stimuli the more rapid is the learning, i.e., the more rapidly the animal learns that *A* leads to *B*. The significance of the relationship between the stimuli at two ends of a short-alley maze should be learned more rapidly than the significance of the relationship between stimuli at the end of a long-alley maze, because the former are spatially and therefore temporally more nearly contiguous.

A few other comments should be made about the perceptual-learning theory. According to this theory, stimuli of any nature which distinguish one experience from another will facilitate learning. Perceptual experiences cannot become discrete unless the stimuli which cause them are discrete. In a sense, this is the point made in our discussions of the rôle of similarity in learning. The more dissimilar are items in a given series the more rapid is the learning. It is likewise true that if different sections of a maze are given distinguishing characteristics, learning will be more rapid than if the sections are relatively homogeneous (similar).

Perceptual learning theorists predict behavior only under known motivating conditions. Without such known motivation, the theorists can only assume that the organism "knows" that *if* a certain stimulus is followed by a certain response (such as running), it will lead to another stimulus. Unless the animal is appropriately motivated, however, such running behavior need not take place. It is presumed that if the animal could perceive the stimulus at *A* in a maze in conjunction with those at *B*, learning would take place even though the rat had never travelled from *A* to *B*.

The theory may be exemplified in human behavior. Suppose we visit a city for the first time. Standing on a corner we look down the street and see the marquee of a theater. According to the perceptual-learning theory, having contiguously experienced stimuli-at-corner and stimuli-at-marquee, we have learned that *if we did* wish to go to the theater (if we were motivated) we could, from that corner, walk a few blocks and be at the theater. Learning has taken place without overtly making the response of walking from corner to theater or being satisfied by so doing.

The theory emphasizes the *place nature* of learning as demonstrated by the rat in the maze. The rat learns, so to speak, that the goal box is "over there"—under that light or close to the radiator. The rat doesn't learn to make a right turn to this stimulus, or a left turn to that one. The rat, in effect, constructs the functional equivalent of a "mental" map of the maze. The locomotion in going to and from various stimuli in the maze is relatively inconsequential to learning; it does, however, allow the stimuli to be experienced contiguously.

Contiguous-Conditioning Theory

This theory has unique principles which set it off clearly from both the theories which we have outlined. It is largely the work of E. R. Guthrie and is found in two of his books (128, 129), both of which are written in such fashion as to contrast sharply with either Tolman's or Hull's presentation. Guthrie uses neither mathematical constructs nor a new system of terminology; he has made neither a formal presentation nor a rigid step-by-step analysis of his theory. In addition to the selected experimental data which Guthrie examines in light of his theory, he leans heavily

on anecdotal evidence. Some of his anecdotes are becoming classic illustrations. For example, here is an illustration of conditioning:

> Two small country boys who lived before the day of the rural use of motor cars had their Friday afternoons made dreary by the regular visit of their pastor, whose horse they were supposed to unharness, groom, feed and water and then harness again on departure. Their gloom was lightened finally by a course of action which one of them conceived. They took to spending the afternoon of the visit re-training the horse. One of them stood behind the horse with a hayfork and periodically shouted "Whoa" and followed this with a sharp jab with the fork. Unfortunately no exact records of this experiment were preserved save that the boys were quite satisfied with the results (128, p. 48).

Guthrie's basic principle is drawn from old associationistic doctrine, but is given restatement in conditioning terminology. The associationistic doctrine is based on the idea that when two events (psychological processes) occur together in time, there is an increase in the probability that one will evoke the other. Contiguity of events thus becomes the basic variable in the learning situation. Temporal contiguity provides the key to learning; *if two events, namely, a stimulus and a response, are contiguous, that is the single sufficient condition for learning.* In the story the stimulus "Whoa" and the response of jumping at the prod of the fork were contiguous. For Hull, the reduction of motivation is the basic condition of learning; for Tolman, contiguity between stimulus events is the prime factor, and for Guthrie, the critical condition is contiguity between stimulus and response.

Another unique aspect of Guthrie's theory is his insistence that the association which is established is an all-or-none affair after a single contiguous occurrence of the two events. If this seems contradictory to the observation that associations are learned gradually, it only appears so on the surface as we shall soon see.

A negative principle which underlies Guthrie's theory is that motivation need not necessarily be present in order that learning may take place. All that is necessary is that a response be produced in some fashion and it will thereupon become connected to the stimulus pattern existing at the time of occurrence. The fact that learning appears to take place only in motivated organisms does not negate the principle, since motivation may and usually does

act as a "forcer" of the response. If the response could be forced in some fashion other than by motivation, learning would take place.

Now, let us restate the essential features of the theory around the four primary variables.

Motivation and consequences. 1. Motivation is not a prerequisite for learning. It is usually present when learning takes place, because motivated organisms are more likely to make responses than non-motivated organisms.

2. The response which is conditioned is a physical response, such as movement of muscles, and no assumption is made as to what the movements accomplish, i.e., whether or not the organism gets a reward. Responses which are evoked just before the receipt of a reward *appear* to be strengthened by the occurrence of the reward only because the reward markedly changes the stimulus complex, thus preventing other responses from becoming associated with the stimulus complex which elicited the response. The stimulus complex includes not only external cues but also intraorganic stimuli resulting from movements and motivational states. The last response made to a given stimulus (recency) is the one which remains associated with that stimulus. The reward changes the stimulus situation, "preserving" it from other responses which occur and therefore replace the response which led to food.

3. Although conditioning takes place in a single contiguous occurrence of a stimulus and a response, we know that some associations are learned more rapidly than others. Responses which are made when the organism is under a state of excitement (strong motivation) appear to become attached to stimuli more quickly than those associations formed under relative quiescence. We must recognize that the stimulus complex is not only a discrete external stimulus such as a bell, a light, or a hayfork. In addition, the stimulus complex is made up of proprioceptive stimuli existing at the time the response occurs. Movement stimuli may become associated with responses, and under strong motivation a great many more of these internal stimuli are so associated. The notion of motivational excitement as a principle of learning is called *vividness* or *intensity*.

Frequency. Learning takes place when a stimulus occurs contiguously with a response; learning is complete with this single conjunction. We observe a gradual rise in the frequency of

response from trial to trial because we are unable to control the stimulus conditions so that the same complex of stimuli is present on each trial. Each time the response is made the stimulus situation is slightly different. Gradual learning *appears to be* observed because each new trial (frequency) increases the number of cues which have become conditioned to the response. The learning curve presumably reaches an asymptote when all possible stimulus components have become conditioned stimuli to the same response.

Contiguity. As we have already mentioned, contiguity of stimulus and response is a necessary and sufficient condition for learning. All other principles actually become secondary.

ILLUSTRATIONS OF EXPERIMENTAL TESTS OF THE THEORIES

None of the three theories we have discussed is comprehensive. All of them have been seriously applied only to simple forms of learning, and in no case has a theory been shown to account for all known facts even at the simple learning level in animals. As we have mentioned before, it is difficult to get tests of the theories at the human level where motivational factors are difficult to control, so the emphasis is on animal experimentation. It must be understood that this is not the place for a comprehensive evaluation of the theories; we shall simply give illustrations of attempts to test the theories. There are no grounds as yet for saying that this theory is no longer tenable whereas that one is, although we may say this about certain single features of a theory.

Tests of the Contiguous-Conditioning Theory

Certain of Guthrie's principles are probably untestable. For example, it would be like chasing a will-o'-the-wisp to test the basic idea that learning is an all-or-none affair with a single conjunction of stimulus and response. No matter how carefully *E* might attempt to reproduce stimulus conditions on a second trial to test the learning which may have occurred on the first, it can always be said when the response doesn't occur on the second trial (and it usually doesn't), that it is because the stimulus situation for the second trial was not exactly the same as for the first.

Few direct tests of Guthrie's theory have been attempted, and as compared with the other two theories reviewed, it has provoked little experimental work. However, certain data are available and pertinent to some of the principles.

The Loucks' study. Guthrie's theory states that all that is necessary for learning to take place is that a response be produced, and the stimuli present at the time will thereby become cues for the response. To test this theory *E* must in some fashion elicit a response which is not followed by reward or punishment, since only by this method can he determine whether learning will take place in the absence of such consequences. Loucks (223) used three dogs in his study. By a minor operation tiny wires were inserted into the motor area of the cortex. The wires were in turn connected to a small coil on the skull so that another induction coil could be used to stimulate electrically (but apparently without causing pain) the motor area of the cortex. The dogs were put through a training period in which the *CS* was a buzzer and the *US* a shock to the cortex, the shock eliciting foreleg flexion. Thus, the response of leg flexion was being forced by electrical stimulation of the cortex. If Guthrie's assumption is correct, after a reasonable number of pairings the buzzer itself should elicit leg flexion.

Loucks gave the animals 600 trials at the rate of 20 trials a day, but at the end of training *no evidence for conditioning was observed,* i.e., the buzzer would not elicit leg flexion. After these 600 trials the procedure was changed slightly, a bit of food being given after each leg flexion. Under these conditions, learning took place rapidly. After the *CR* had been established, the food reward was removed, and the response promptly extinguished. Thus, Loucks' study gave no evidence of learning when a *CS* was contiguous with a response unless that response was followed by reward. Therefore his results do not support a contiguous conditioning theory. Comparable results have been found in another similar study (224).

The Seward test. Guthrie holds that the last response made to a stimulus is the response learned. All other responses which were previously associated to the stimuli are dissociated when the new response occurs. This is a principle of recency, and Seward (346) sought a test of it. He reasoned that if an animal, after making a

response, was immediately taken out of the situation and then later put back in, the response should again be made, since it was the last one elicited by the situation. Rats were placed in a Skinner-box, and as soon as they made a lever-pressing response they were taken out of the box. Members of another group obtained a pellet of food for pressing the lever before they were removed. A control group was neither removed from the box nor given food for pressing the lever. Each rat of the two experimental groups was given 20 trials, a trial being one lever-pressing response followed by removal. The animals of the control group merely spent 1 minute in the box each day.

Seward's results show that the rats which were removed from the box immediately after pressing the lever did learn, as measured by a decrease in time, to press the lever. However, the amount of learning was quite small as compared with the group given food before removal. Extinction trials, in which the animals were left in the box for 7 minutes for 3 consecutive days, show that the removed group made more responses than did the control group. Thus, there is some evidence for this particular aspect of the Guthrie theory, and while the learning which did take place by removal may be quite inefficient as compared with rewarded learning, some learning did occur. It may be, of course, that removal from the box had some reward value for the animals, although it is difficult to specify what that is, since considerable precaution was taken against secondary rewards being operative and against the rats receiving food for some time after removal.

In an additional experiment (347) Seward and his students took a different approach to the problem. According to the contiguous-conditioning theory, the last response to a stimulus is the only one which remains attached to the stimulus. If this is true, and if two responses are attached successively to the same stimulus, only the second one should be given later to that stimulus. Seward required human *S*s to learn two motor responses to the same stimulus and then tested to see if only the second occurred. He found that the first learned response occurred just as frequently as the second. There are many data from verbal learning which would support this finding. It should be pointed out, however, that such a result does not negate the Guthrie assumption. Although the second stimulus may appear to be objectively the same as the

first, there are probably other factors by which one might distinguish between the two. Learning the second response does not necessarily remove the first response, since it may have been given to a slightly different stimulus pattern. Since we might say that the basic theory could not possibly be tested, we must therefore assert that the above results do not negate the theory as stated.

Guthrie's own work. Several of Guthrie's assumptions, especially those concerned with protection from dissociation of the last response made to a situation, have been derived from his experimental work with cats in puzzle boxes (131). On a series of trials the cat was placed in the box, and when it hit a pole the door opened, allowing access to a bit of salmon. Few quantitative data have resulted from this work; rather, a series of pictures of the cat's movements in getting out of the box is presented. Guthrie has placed considerable emphasis on movement stimuli, that is, stimuli produced by the organism's own movements which serve as stimuli to successive movement responses. Movements and postures make up the behavior which Guthrie is primarily interested in predicting, not how soon the correct response is hit upon or what goal the correct response achieves for the animal.

To appreciate fully the evidence on which Guthrie bases this theory, it would be necessary to observe the movies or the many drawings of successive postures of the cats in the box. The best we may do, however, is to quote his basic statement concerning these studies:

> We may summarize our generalizations of the puzzle-box behavior as follows: The behavior of the cat on one occasion tends to be repeated on the next, even to occasional prolonged series of movements about the cage. Exceptions to this are either the results of a different entrance, which initiates a different line of action, or the result of accidental distractions, which may deflect the behavior from its former sequence; moreover, when the cat has been in the box for a long time, responses made later in the trial may supplant those made earlier.
>
> The most stable response is the one which ends in release from the box. In some instances this is a comparatively long series of movements ending in exit. The reason for its stability appears to be that escape removes the animal from the puzzle-box situation so that no reassociation may occur. All response series tend to recur, but some of them are lost when the animal remains in the box and is compelled to new action. *The cat*

> *learns to escape in one trial* and will repeat the specific movements of its first escape except is so far as new trials by accidental variations of situation cause new associative connections to be established. A protracted trial breaks up the sequence of movements by superimposing new behavior in the same situation. Time, therefore, tends to be reduced with repetition (131, p. 41).

It is not, of course, inevitable that the empirical facts upon which Guthrie bases the generalizations must be explained only by a contiguity-conditioning theory. Other theories can adequately embrace the findings. Yet, the relative simplicity of the theory has won it many adherents, and we shall very likely find it intriguing experimentalists for some time to come. No more will be said here about the contiguous-conditioning theory. The experiments to follow are concerned with the perceptual-learning and goal-reinforcement theories.

Latent Learning

The studies we shall examine under this heading probably represent as near "crucial" tests of opposed theoretical points of view toward learning as it is possible to devise at present. The opposed theories in question are the perceptual-learning and goal-reinforcement theories. The word *crucial* has been placed in quotation marks because the data as they stand today are contradictory; one set of data entirely supports one theory, whereas a different set just as strongly favors the other. The nature of the experiments from which each set of data has been derived is somewhat different, but both are presumed to test the same basic proposition. Let us review this proposition.

The perceptual-learning theory assumes that learning consists in the perception of relationships between stimuli. Simplifying to it baldest terms, if an organism experiences contiguously one set of stimuli, A, with another set B, which includes food, it will on later occasions expect B when A is present. The goal-reinforcement theory assumes that in order for stimulus A to be followed by a response which leads to B, A and the response must have been followed previously by motive reduction as a result of eating food at B.

It appears that here is a clear-cut difference in theory which can be put to crucial experimental test. If a rat goes from A to B

and gets no food, or does not eat food at B because he is satiated, learning *should* take place according to the perceptual learning theory and *should not* take place according to the goal-reinforcement theory. Experiments of this nature are called *latent learning* experiments. Latent learning would be learning which, in so far as E can tell, has taken place without reward being present. First, we shall consider studies which support the perceptual-learning theory and then studies which support the goal-reinforcement theory.

Buxton's experiment. Buxton (53) used several groups of rats in his experiment, but we shall consider procedures for only two groups since they will provide us with adequate knowledge of methods and results. A multiple-T maze with 12 choice points was the basis for the learning problem. A group of male rats was placed in the maze on three consecutive nights for about 16 hours each night. None of the rats had been fed and watered for several hours before being placed in the maze, so we may assume that they were quite hungry and thirsty before the night was over. The maze had the usual starting box and food box, *but at no time during the three nights was food or water present in the maze.* During the three nights the rats did, of course, explore the maze thoroughly, and descriptively we might say they had opportunity to get acquainted with it.

Following the three nights' experience, each rat was individually placed directly in the food box, then containing food, and after being allowed to eat for about 30 seconds, was removed from the food box and placed in the starting box. The question was: "Did the rats learn anything about the maze during the three rewardless nights?"

To answer this question, it was necessary to have a control group which had had no initial experience in the maze. This group consisted of animals which spent three nights in a straight-alley maze located in the same room as the multiple-T maze. Placing the straight-alley maze in the same room gave the control group the same opportunity to "get used to" the same noises, smells, lights, and objects outside the maze to which the experimental group was exposed. Thus the control group had exactly the same experiences as the experimental group except for multiple-T maze exploration. The control group was tested in

the multiple-T maze in the same fashion as the experimental group; first they were allowed to eat food in the food box and then they were removed to the starting box for their first run.

If latent learning took place, we would expect the performance of the control group to be poorer than that of the experimental group. The first trial is the critical one. The results, in terms of number of errors made at each choice point for the two groups on the first trial, are shown in Fig. 76. Two facts stand out in

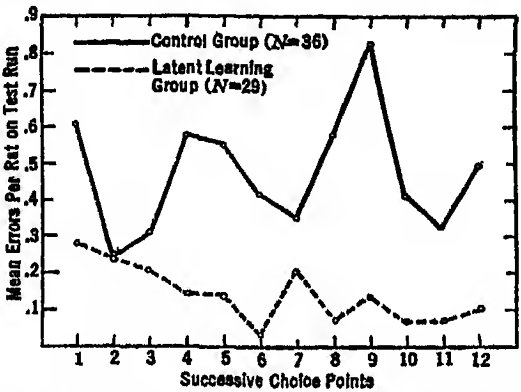


Fig. 76. Latent learning in a multiple-T maze. Data from Buxton (53).

this graph: (1) the total number of errors made by the control group is considerably greater than the total made by the experimental group, and (2) the number of errors made at each successive choice point by the experimental group gradually reduces in frequency as we go from the first to the twelfth. This latter finding is indicative of the goal gradient with which we are already familiar (Chapter XI), for it shows that fewer errors were made by entering blinds close to the goal box than were made by entering blinds immediately after leaving the starting box. Since this phenomenon appears on the first rewarded trial, we must infer that learning had taken place during the nights the rats spent in the maze.

Another aspect of the goal gradient is that the speed of locomotion of the rat increases as it runs from the beginning to the end of the maze. This evidence of learning was present on the first rewarded trial for the experimental group, but not for the control

group. Thus, we have further evidence from which we must infer that learning had taken place before the first rewarded trial.

On the surface these results may appear quite reasonable. Why shouldn't the rats that had spent three nights in the maze know how to get around it? Why shouldn't they run faster than the control group which had never been in the maze before? From the goal-reinforcement point of view, however, explanation of these results is not simple. The goal-reinforcement theorist must ask: "Where is the reward or the reinforcement which occurred during the nights the rats stayed in the maze?" According to this theory there must be motive reduction or symbolic reward of some kind before learning will take place. Certainly the rats were motivated—they were hungry and thirsty—but they got neither food nor water, and there is no reason to believe that secondary reinforcement of any kind was present. These are the puzzling problems which the latent learning experiments of this type have raised. Let us consider another method of demonstrating latent learning before pursuing this problem further.

Blodgett's experiment. Blodgett (28) used a six-choice multiple-T maze and three groups of rats. The control group was given food after each trip through the maze. Another group went through the maze twice without being fed, and then on the third trial and every trial thereafter they were fed in the food box as were the members of the control group. A third group was given six runs without being fed and then on the seventh trial and every trial thereafter was fed. All groups were given one run per day for nine consecutive days.

The mean errors of the first group, the group rewarded after each run, show the expected reduction in frequency from trial to trial for the nine trials. The second group showed a slight reduction in mean frequency of errors on the first three runs, but on the fourth run a rather large reduction in number of errors took place after having been rewarded on the third. The most striking results were shown by the third group. For seven trials there was only slight reduction in the number of errors. At the end of seven runs the animals were still averaging well over two errors per run (three would be chance in a six-choice maze), whereas the other two groups were averaging about .25 errors per run. However, on the eighth run, after having been rewarded

on the seventh, the third group showed a striking reduction in the number of errors, and by the end of the ninth run this group was performing as well as the other two groups. *It would appear that learning had taken place during the first six runs even though it was not demonstrated in the error measurement on those days. When food was given, the latent learning showed itself.*

Blodgett's and Buxton's findings, taken in conjunction with the results of other similar experiments, are strong evidence for the perceptual-learning theory. Learning appears to take place without reward; all the organism has to do is experience stimuli contiguously. According to a goal-reinforcement theory, no learning should have taken place on the rewardless runs of these experiments. Goal-reinforcement theorists have not as yet offered a satisfactory explanation of these findings. In support of such a theory we might say that there *must have been* rewards which were "hidden" or secondary in nature. If we did offer this "explanation," however, we would be inviting the serious criticism that such a position is untestable. For, we would in effect be saying that we do not have to identify rewards, i.e., we merely beg the question by assuming they are present when learning occurs. This, of course, we cannot do. On the other hand, it would be good scientific procedure to state an hypothesis as to the nature of the reward (or punishment) in the latent learning situation and then proceed with an experimental test of this hypothesis. Since this has not been done, we must conclude that the goal-reinforcement theory does not yet explain the results of the above experiments.

Let us turn to other experiments which provoke the same kind of unrest and doubt in the perceptual-learning theorists as do the above experiments in the goal-reinforcement theorists.

The Spence-Lippitt experiment. Spence and Lippitt (364) used a simple Y-maze as shown in Fig. 77. During training all rats were satiated for food but were thirsty for water as a consequence of 18-hour water deprivation. The right alley of the maze always led to water for all animals. For 10 animals, the left alley led to food whereas for 10 others the left alley led to nothing except the end box. In no case did the animals of the first group eat the food in the left end box, since they were thoroughly satiated. However, when the animals in both groups

took the right alley they were rewarded by being allowed to drink water. Both groups of rats were given 5 trials a day for 12 days.

Let us note the rationale of this experiment. When any animal took the alley to the right, it would be rewarded with a drink of water. When the members of one group of 10 animals went down the left alley and saw the food, they would not eat because they were satiated. However, the stimuli of the left alley, followed by the experience of seeing food, should produce learning to the effect that "left alley leads to food" according to the perceptual-learning theory. The 10 animals which found no food at the end of the left alley should also learn, according to a perceptual-learning theory, but they would learn only that "left alley leads to box." After training had taken place in both alleys, *E*s had to determine whether or not such learning as predicted by the perceptual-learning theory actually took place.

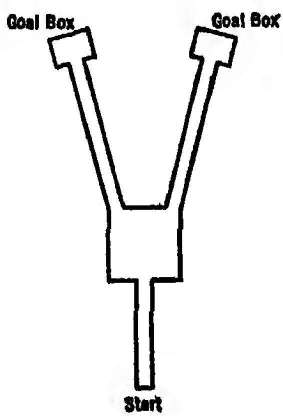


FIG. 77. Simple Y-maze ed by Spence and Lippitt 64).

A methodological difficulty which Spence and Lippitt faced in this experiment should be discussed. Regardless of the theory to which one adheres, motivation is held to be an important determinant of responses. By any theory one would predict that the rats in the present experiment would, when satiated for food and thirsty for water, learn quickly to take the path leading to water. Such being the case the animals would get overwhelming experience of the alley leading to water as compared with the other. In order to give both groups of animals experience in going down the left alley, the water alley was blocked half the time. By forcing the animals to run to the left half the time, both alleys were experienced equally. As an added precaution, the door leading to the left alley was sometimes closed so that the animals would experience runs to the right alley with the left alley blocked, as well as runs to the left with the right alley blocked.

The critical point of the experiment came on test trials when

the animals *were satiated for water but hungry for food.* Thus, the motivating conditions were just the opposite of those which prevailed during training. What are the differences in the predictions which would be made by the two theories on the test runs? The goal-reinforcement theory would say that the rats in both groups would take the alley to the right—they would execute the response for which previously they had been rewarded. The perceptual-learning theory would predict that the rats which had seen food at the end of the left alley on the forced runs would now take that alley when hungry; seeing food in perceptual contiguity with the stimuli at the choice point would be enough to provide learning. It would not be predicted, of course, that the group of rats which had seen only the empty end box would take the left path.

The results show that on the first test trial all animals of both groups took the right path—the path leading to water. These results are exactly what would be predicted by the goal-reinforcement theory and quite the opposite of the prediction by the perceptual-learning theory. Furthermore, both groups were given additional learning trials under the food-motivated condition. The perceptual learning theorist would say that at least the group having had experience in going down the left alley and perceiving food would learn this response most rapidly. The results show that both groups learned this habit in about the same length of time, so this prediction is not substantiated. These findings heavily favor a goal-reinforcement theory and are quite contrary to predictions of a perceptual-learning theory.

Three additional experiments of the general nature of the Spence-Lippitt study have been performed and all three support the original findings (123, 194, 395). In these later experiments special precautions were taken to assure that during training trials the satiated rats saw the food even though they did not eat it. In one of the studies (123) the rat took either of the two alleys to get water but there was food in only one goal box, this food being scattered profusely over the floor of the box so that the animal literally had to wade through it to get to the water bottle. Still, on test trials, when the animals were hungry but not thirsty, no evidence was obtained indicating that the animal had learned anything about how to get to food.

One incidental finding of these studies is that the sight and smell of food does not seem to provide any appreciable secondary reinforcement when the animal is motivated for water. In view of the fact that in home cages food is typically present when the animal drinks water, such secondary reinforcing properties might be expected. Since this situation is conducive to the development of symbolic rewards, we might assume that the left alley response would be strengthened somewhat as a consequence of the rat's seeing food (see Brogden's experiment in Chapter VI). That there was no difference in the performance of the two groups of animals used by Spence and Lippitt indicates that secondary reinforcing property of food in the left alley was of small consequence in this situation.

One further set of data should be mentioned. Meehl and MacCorquodale (251) carried the Spence-Lippitt procedure one step further. Rats were placed in a simple T-maze while satiated for *both* food and water. Food was placed at the end of one arm of the maze and water at the other. Although the "running" was slow because of lack of a dominant motive, 4 trials a day were completed for 10 days. Half the runs were to the water end and half to the food end, the runs being forced when necessary in order to maintain the 50:50 ratio. The 24 animals thus had experience in going to food (but not eating) 20 times and going to water (but not drinking) 20 times. Following this training, half the rats were satiated for water and made hungry for food and the other half were made thirsty for water while satiated for food. Did the training experience produce learning so that the animals would now make the response leading to the incentive appropriate to the motivating condition?

The results show that some learning had taken place since the animals responded correctly on the test trial more frequently than would be expected by chance. It appeared as if the animals had learned merely by seeing the incentives on the training trials. However, subsequent runs showed that the learning was very weak as compared with that which took place when actual reinforcement by eating or drinking occurred. Furthermore, Meehl and MacCorquodale offer a plausible explanation of the learning on the 40 trials in terms of secondary reinforcement and feel that it is unnecessary to invoke perceptual learning to account for the

findings. Their explanation is too complex to justify inclusion here.

What are we to make of these two sets of experiments, one set favoring one theory, the other set favoring another? There seems to be no answer to this question at present. The findings of the experiments have been so clear-cut that the burden of explanation rests entirely on interpretation and not on questioning the validity of the findings. We may note that the Spence-Lippitt experiment provided primary reinforcement of one response at the same time that latent learning was supposed to be taking place for a different response. This was not true in the Buxton or the Blodgett experiments. The learning in the Buxton and Blodgett experiments took place when the animals were motivated but when there was no incentive. The learning took place in the Spence-Lippitt experiment when the animals were motivated for water and received water incentive, but failed to take place when the rats were not motivated for food although the food incentive was present. The rat was given much more freedom of discovery (and the maze was more complex) in the Buxton and Blodgett procedure than it was in the Spence-Lippitt method. It is difficult at present, however, to proceed from the recording of such differences to account for the obtained results. Apparently we must at present conclude that the two sets of data are contradictory as far as the theories go. Probably we may also conclude that this situation will be quite provocative of additional research.

Place Learning

A demonstration of place learning. We have seen that the perceptual-learning theory emphasizes the directional and place nature of any learning which requires locomotion through space. When an animal learns to go through a maze, it does not learn merely to make a right turn here, or a left turn there, but rather tends to learn that in relation to its present position in the starting box the goal box is "over there." The locomotion and turning are merely means to an end; what is learned is the spatial relationships among various path stimuli, stimuli outside the maze, and goal-box stimuli.

It is fairly well recognized that maze learning may involve a great many cues outside the maze, such as lights, walls, and other

constant stimuli which can be perceived by the animal. These stimuli, according to the theory, allow the rat to build up place knowledge and learn that the goal box is under that light or close to the window. Such inferences can only be drawn by manipulat-

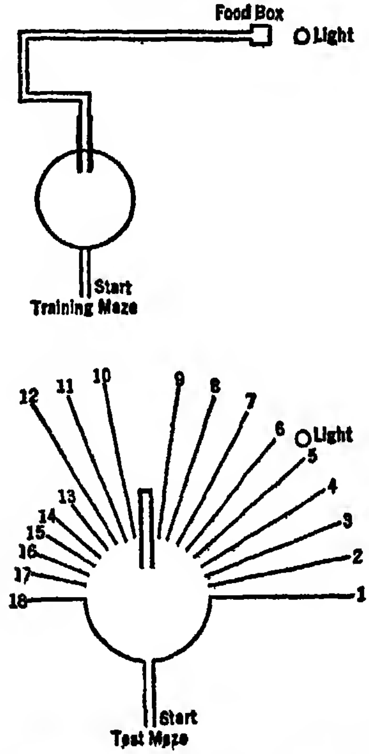


FIG. 78. Elevated mazes used by Tolman, Ritchie, and Kalish (387) in their demonstration of place learning. The fan-like lines on the test maze are intended to represent alternative paths similar to the one shown with both walls.

ing conditions under which a rat learns a maze until these inferences seem plausible. During the last few years there have been several studies in which just such manipulations have taken place. We will sample these studies. First, let us cite one of a series of studies by Tolman and his students (387) which demonstrates what is meant by place learning.

To understand the demonstration we must understand the apparatus. It consists of two parts, the training maze and the test maze. The training maze (elevated) is shown in the upper part of Fig. 78. The only part of the maze that was enclosed is that portion leading from the round table. Note the light (a five-watt bulb) which was placed near the goal box as the only source of illumination in the room. Initially the rats were trained to run from the starting point to the goal box. This was largely a matter of getting the rat adapted to the situation, since there are no blind alleys. Adaptation was gradual, in that initially the rat was fed in the goal box, then given short runs to the goal box, and finally given five complete runs through the total length of the maze.

The test maze is shown in the lower portion of Fig. 78. All the alley beyond the circular portion has been removed except the sided portion leading from the table. A radial series of 18 alleys now leads off from the circle. All these alleys are 6 feet long, except alleys 13 to 18 which were restricted because of the size of the room. The alley which the animal originally took was blocked. The question concerns the alley which the rat will choose to go through on its first and only test trial.

Three of the 56 animals made no choice within 6 minutes and were therefore discarded. Typical behavior of the other rats involved going first into the blocked alley, then returning, exploring the round portion, and tentatively entering at least one other alley before entering one completely. If a rat took one of the six short alleys, it was allowed to come back to the table top to choose another. The test trial was considered completed when an animal reached the end of one of the 12 long alleys. The per cent of rats taking each alley is shown in Fig. 79. Alley six was the alley which pointed most directly to the point where the food box had been previously, and clearly, as compared with other alleys, this alley was taken with greatest frequency (36 per cent). It is the shortest path to the goal which was not there. Let us quote *Es*' conclusion:

The fact that they (the rats) selected the shortest path indicates that what was learned during the preliminary training was not a mere response sequence, or an expectation that this particular path led to the goal. They learned, instead, a disposition to orient towards the physical location

of the goal. Because of this we have chosen the word "expectation" as the name for this orientational disposition (387, p. 23).

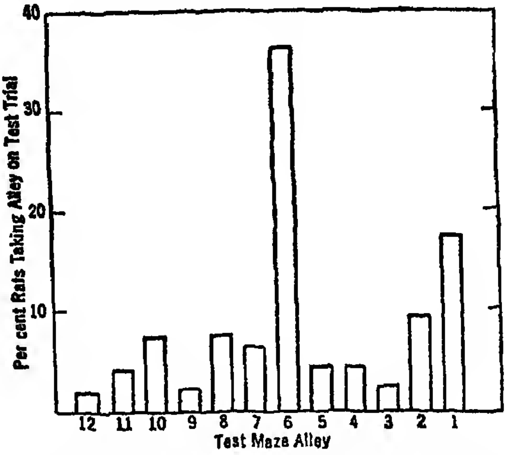


FIG. 79. *A demonstration of place learning. Data from Tolman, Ritchie, and Kalish* (387).

Place learning vs. response learning. The above quotation indicates these adherents of place learning theory believe that learning is much less specific than is implied by a person who says that a rat learns to make this particular response at this point, another at the next point, and so forth. In pursuit of a test of these two opposed interpretations, Tolman, Ritchie, and Kalish (388) devised a situation which they believed would separate place learning from response learning. The maze pictured in Fig. 80 was the basis for the experiment. The two food boxes, F1 and F2, could be reached from two starting boxes, S1 and S2. When the rat was placed in S1 the response rewarded was a right turn; if the animals took the path to F2 they found no food. If the animals were placed at S2, the rewarded response was that of taking the F2 alley, F1 yielding no food. Note that irrespective of the starting box in which the rats were placed, *the correct response was a right turn at the choice point.* These conditions held for the response-learning group of animals.

A second group of rats was trained as a place-learning group. For half the animals in this group the rewarded response was

going to F1, whether they were started at S1 or S2. The other half of the animals was rewarded at F2, whether started at S1 or S2. In both groups of place-learning animals, the correct response required a right turn half the time and a left turn half the time, but always, whether right or left, it required going in a constant direction. In the second group the emphasis was upon getting to a certain place whereas in the first group the emphasis was upon making right turns.

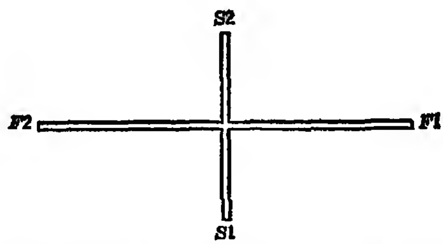


FIG. 80. Elevated maze as used by Tolman, Ritchie, and Kalish in their study of place learning vs. response learning (388). *S1* and *S2* indicate starting points; *F1* and *F2* indicate food boxes.

The first group, the right-turning group, was run for 12 days with 6 trials per day. Out of eight rats only three reached a criterion of ten successive errorless trials. The other five rats developed a tendency to go always to the same place—F1 or F2—and as a consequence they continued to make 50 per cent errors since the "place" was correct only half the time. In contrast, the second group made no further errors after an average of only 3.5 trials —they rapidly learned to go to the correct place. Thus, the results would seem to favor a place-learning theory rather than a specific turning-response theory.

The Blodgett-McCutchan interpretation. Blodgett and McCutchan (29) questioned the interpretation which Tolman, Ritchie, and Kalish gave their results. Blodgett and McCutchan believed it was possible that the so-called response-learning group actually might have been learning two alternate place responses, one to go to F1, the other to F2. This would make the task more difficult than learning a single place response, as was required of the second group. In short, Blodgett and McCutchan did not believe Tolman and his students had clearly separated response

learning from place learning. The experimental conditions which we shall report now tend to support the validity of their disbelief.

Blodgett and McCutchan arranged four conditions, using a separate group of rats for each condition. The first condition was designed to make response learning a necessity if the animal was to learn. The apparatus was a simple T-maze, set under a circular dome. The walls of the dome were homogeneous. A single lamp was placed in the center of the dome, and a rheostat permitted dimming the bulb until the room was almost dark. To run the trials it was first necessary that *E* remain in the room for 15 to 20 minutes to get dark-adapted. When so adapted, *E* could barely see the outline of the rat. Under such conditions it was supposed that the rat could get practically no extra-maze cues. Furthermore, the position of the maze was changed from trial to trial so that even the very dim light could not have aided the rat. The animal, functionally speaking, was blind. Furthermore sounds were at a minimum and odors were random. If the rat learned, it would learn largely through proprioceptive and touch cues and not through place stimuli mediated by distance receptors.

The trials were given by having an assistant pass the rats to *E* through a door and record errors and correct responses as reported by *E* after each trial. Each rat was trained to make a specific turning response (right or left) by rewarding that response only. Animals were tested initially, in order to determine the turning preference of each, and on learning trials were rewarded for making opposite turns. Thus, if an animal showed an initial right turning bias, during the training he was rewarded for making left turns in the maze. Let us call these animals Group I.

Group II also learned in the T-maze set under the dome. However, the light on the dome was on at its full wattage (75 watts) and the maze, placed to one side of the light, always remained in a constant position. Under these conditions, the animals could learn either a place habit (toward or away from light) or a specific response (right or left turn).

Two additional groups of rats were run in another room under conditions very similar to those used by Tolman, Ritchie, and Kalish, and the results of these *E*s were substantiated. However,

the important finding of the Blodgett-McCutchan experiment was that Groups I and II learned equally well (Fig. 81). The conditions of Group I were such that the animals in that group must necessarily have learned responses to proprioceptive and touch cues. The members of Group II might have learned by such cues, or by the place location of the goal box, i.e., its relation to the light. The fact that the members of this group learned no more

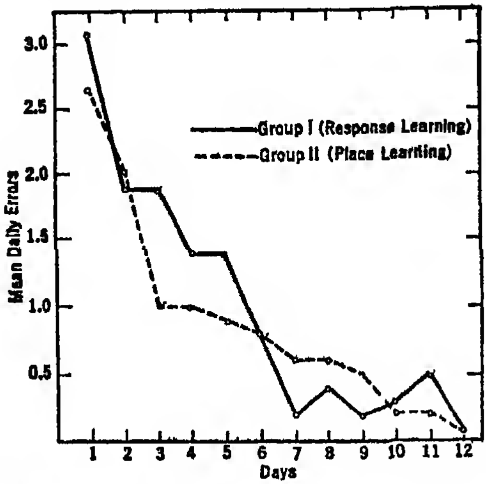


FIG. 81. Place learning vs. response learning in a simple T-maze. Estimated from Blodgett and McCutchan (29).

rapidly than did members of Group I suggests that place learning, if used by members of Group II, was no more efficient than response learning, and if not used, it shows that rats did not necessarily "prefer" place learning.

These experiments on place and response learning are sufficient to show us the differences in points of view and to demonstrate methods which have been used to test those differences. The findings imply that no critical results in favor of one theory over another will be found by such experiments. Furthermore, there is reason to believe that the predictions of a goal-reinforcement theory and those of a perceptual-learning theory would not necessarily differ on these matters. A goal-reinforcement theory does not imply that responses cannot be learned to stimuli outside the

maze as long as responses to those stimuli are rewarded. If an approach response to a light stimulus is rewarded by food in a box near the light, that response (approaching light stimulus) gains strength. Similarly, there is no reason why a perceptual learning theory cannot assert that the animal learns the significance of proprioceptive stimuli to the food stimuli. In short, it does not appear likely that place vs. response learning is going to provide critical data for theories in the sense that the latent learning studies have provided such data.

The Spread of Effect

This phenomenon known as *spread of effect* will have to be treated briefly and without arriving at definite conclusions. We cannot, however, pass it by in even a brief treatment of theories of learning. The fact that we shall be unable to cover adequately the numerous ramifications of the spread of effect speaks for the number of varied experiments which have been performed on it.

The spread of effect is one of Thorndike's (379) discoveries. Let us take a simplified illustration of the operations which are necessary to define it. As *S*, you are instructed as follows:

> I am going to read you a series of words. To each word I have assigned a number between one and ten. After I read each word you are to guess a number between one and ten. If your guess is correct I will say "Right"; if your guess is incorrect I will say "Wrong." I will read through the entire series of words, and then repeat it. Your task is to learn as many of the correct numbers as possible.

E, by a pre-arranged plan, has "rights" and "wrongs" arbitrarily assigned to words regardless of what *S* says. Let us say that we have nine words, and *E* has determined beforehand that he is going to say "right" or "wrong" to each word as indicated:

Stimulus Words	*E's Responses*
JAY	wrong
SOD	wrong
RUM	wrong
BET	wrong
CAP	*right*
KID	wrong
TUG	wrong
SON	wrong
SAW	wrong

Regardless of what number *S* gives as a response to the stimulus *JAY*, *E*'s response will be "wrong." Regardless of what response is given to *CAP*, *E*'s response will be "right." *E* records each response given by *S*. On the second trial the same procedure is followed, except, of course, the response which was given the first time to the stimulus *CAP* remains the correct response to that word.

Two trials are enough to define spread of effect. We ask the question: "On the second trial what was the per cent of repetition of each response which was wrong on the first trial, and what was the per cent of repetition of the right response?" If a great many *S*s are used, the mean percentages become fairly stable, and it will be found that the per cent of repetition of the response to *CAP* is the highest of all. This is as we would expect, since we have accepted the fact that rewards (by some mechanism) enhance performance. The data on the repetition of the "right" response presumably demonstrate the operation of the law of effect. But now, we calculate the per cent of repetition of the wrong or punished responses, and we discover that the punished responses next to the rewarded responses are repeated more often than chance would allow. Furthermore, we find that the more remote a punished response is from a rewarded response, the closer the frequency of repetition is to chance. This gradient of repetition of punished responses around a rewarded response is the spread of the effect.

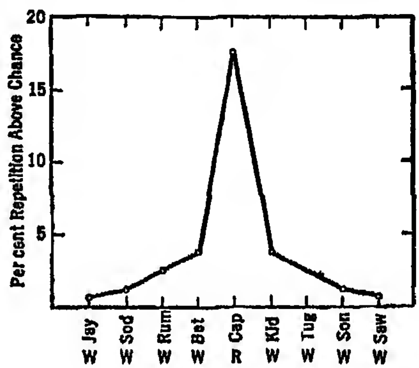


FIG. 82. Representation of spread of effect. W indicates that *E* responded "Wrong"; R indicates that *E* responded "Right."

Using the nine stimuli above, we may graph the expected finding as in Fig. 82. The figures along the ordinate are hypothetical but represent a fairly characteristic frequency of repetition for each position. The gradient is drawn symmetrically, although

usually the repetition of punished responses following the rewarded response is somewhat higher than for those preceding it. It may be noted that the frequency of repetition above chance is not great, a matter of a few percentage points, but the fact that such gradients have been found rather consistently leaves no doubt as to their validity.

We have simplified the situation somewhat. Usually *E* will use many more words, at least 20 and more, probably 60 or 70, and repeat the pattern of a "right" flanked by four "wrongs" several times. The words will be read at the rate of one each 2 or 3 seconds and of course it is very difficult for *S* to attempt to remember specific responses from one trial to the next. Also, in most of the experiments the series has been repeated more than twice, but the computation and interpretation are much simpler if *E* gives only two presentations.

What is the significance of the spread of effect for learning? It was Thorndike's belief that the spread of effect represents the automatic action of reward, not only in strengthening the response which it follows but also in strengthening any responses which were contiguous to it. By making the list long, *S* could not remember the wrong response given to a stimulus on the preceding trial. The fact that this response tended to be repeated when contiguous to a rewarded response suggested the automatic influence of the reward.

As previously indicated, there can be no doubt as to the authenticity of the finding under the conditions which Thorndike and others have used. However, several *E*s have challenged Thorndike's interpretation of the finding and have proceeded to show how the phenomenon is influenced by many factors and how it might be explained by other means. They do not seriously challenge the strengthening effect of the reward on the rewarded connection, but they do question the interpretation which says that the frequency of punished repetition is to be attributed directly to the reward. Let us mention just a few of these findings. If *S*s are not instructed to learn or remember correct responses, there will be no strengthening of rewarded or punished contiguous responses (396). The occurrence or non-occurrence of the spread is a function of the degree of similarity of the rewarded items with contiguous items; the less the similarity of rewarded and

flanking items, the greater is the spread (427). Some evidence has been found to show that at least a portion of the forward gradient may be explained on the basis of number-series biases (181). Thus, if the rewarded response were *four*, *S* was more likely to say certain numbers following four than he would others. Perhaps 4-621 was *S*'s favorite telephone number or part of the sequence of his army serial number. In any event, we cannot assume that each number is a product of chance frequency. Thus, there is a problem concerning the true base line above which the repetition can be said to be greater than would be expected without the rewards and punishments.

In drawing Fig. 82 we mentioned nothing concerning the baseline which should be used to calculate frequency of repetition above chance. What should this be? We know that it cannot be a sheer chance basis, that is, one in ten if there are ten numbers. Number biases rule this out. Furthermore, *S*s are probably unlikely to repeat a number given for the immediately preceding stimulus. It seems clear that the baseline must be established on an empirical basis; we might use a control group or control condition in which we discover the frequency of repetition from trial to trial without *E*'s responding "right" or "wrong." Then, under a condition on which *E* does respond "right" and "wrong," we may discover how much greater the frequency of repetition is above "empirical" chance.

But is this the correct baseline? That is, is it correct to discover the expected repetition by *E*'s not responding at all? Probably not. If we wish to discover the influence of a "right" on contiguous punished responses, our best estimate of frequency of repetition without the "right" would be determined by *E*'s saying "wrong" to all responses. This, compared with repetition when one in nine responses is rewarded ("right"), would give us the estimate of the influence of the "right" on contiguous punished responses.

We should also mention that since the words are repeated in the same order on the two or more trials which are given, serial-position effects (the bowed serial-position curve) would also be expected to have some influence on frequency of repetition. If the baseline is determined empirically, this is taken into account automatically.

Is there an effect produced by saying "right"? Yes. Thorndike's long series of experiments leaves no doubt that such rewards tend to strengthen correct responses, although we cannot say that such rewards will inevitably produce strengthening. Is there direct spread of effect to contiguous punished (or non-rewarded) connections? The evidence is inconclusive.

The Rôle of Frequency

When we speak of the rôle of sheer frequency, we are again speaking of the problem of learning in the absence of rewards and punishments. We have seen that some of the latent learning experiments suggested that learning seems to take place merely by repetition of a behavior sequence. But other experiments which we reviewed suggested that learning would not take place merely as a consequence of frequency. Let us briefly examine data on human learning. We may turn to Thorndike's work for an illustrative experiment (378, Exp. 5).

S drew lines of specified lengths with no information as to accuracy and without ever seeing the lines he drew. He attempted to draw lines 2, 4, 6, and 8 inches long and made 950 attempts for each length. A study of the length of lines actually drawn showed that there were gradual shifts in the length of these lines from time to time. If on the first 100 attempts to draw a line 2 inches long the mean of the lines actually drawn was $2^1/_4$ inches, this result did not mean that $2^1/_4$ inches would therefore be the line length most likely to be drawn on the next 100 tries. Thus, even though drawing a $2^1/_4$ inch line was the most frequent response, that response was not fixated as a consequence of the great frequency given it.

With humans it is not easy to arrange a situation in which repetition of a stimulus-response sequence is provided without consequences or, rather, with a neutral consequence. Thorndike's many experiments have probably approached the criterion of neutral consequences as nearly as possible, and his results have consistently shown little or no increase in the strength of associative connection as a result of sheer frequency of repetition.

The Rôle of Contiguity

We have seen that theories of learning differ regarding the events which are held to be of necessity contiguous in time, in order that learning will take place. Conceptually, contiguity must be taken to mean simultaneity, in that two events, to become associated, must overlap. When we increase the time interval between stimulus and response (or between response and consequence), we do not change our conception of overlapping events. From the fact that two events do not necessarily have to overlap objectively to become associated, we infer that the first has left a *trace*. Trace is a theoretical construct and need not be thought of as physiological in nature. By increasing the time interval between the first and the second event we presume to reduce the strength of the trace of the first when it overlaps the second. When a trace "dies out" so that it does not overlap a succeeding event, the two obviously cannot become linked in causal relationship. When an interval between events is of such length that no learning takes place, we presume that there has been no overlapping of pertinent events or traces thereof.

Thus contiguity remains largely an empirical problem of demonstrating the limits of learning and gradients of efficiency of learning as a function of (1) the interval between response and reward, and (2) the interval between stimulus and response. To date this work has been almost entirely concentrated at the simpler levels of learning. The symbolic capacity of human adults and the operation of secondary rewards make it a very difficult problem to attack in a precise quantitative manner at the level of complex learning.

A SUMMARY OF CENTRAL PROBLEMS OF FUTURE RESEARCH

At various points throughout the last four chapters we have pointed out problems which seem quite likely to engage *Es* in the near future. Let us summarize these problems.

1. We are just beginning to realize the importance of secondary reinforcement brought about by symbolic rewards; symbolic punishment no doubt also occurs. Suppose we built a secondary

reward value "into" a food box. Will that food box serve as a reward for a rat satiated for food but thirsty for water? How *specific* to a given motive are secondary rewards? Are secondary rewards and punishments always of less consequence than primary rewards in animals? In humans? Symbolic rewards in animals may account for a great deal of behavior which is not at present easily interpreted.

2. The effects of punishment on learning still remain ambiguous. Thorndike's work (380) has led him to conclude that punishment does not directly weaken responses which it follows, but will tend to make for varied behavior, thus increasing the probabilities of hitting upon the correct response. Some evidence we have examined, (Fixations, Chapter VIII), suggests that punishment properly manipulated *may* actually strengthen responses. There is also some evidence that punishment may strengthen responses in a relatively non-traumatic situation (cf., Stone, 368). The effect of punishment on learning must in part be a function of the particular situation and what the organism is able to do in the situation when punished, but we have no clear-cut evaluation of differences in these situations. Human adult responses may perseverate even though punished, as seen in some neurotics, criminals, alcoholics, and sex perverts. We may say that there are "really" rewards in these cases and there may well be, but that only makes the problem of factoring out the relative effects of multiple consequences (both rewards and punishments) more difficult. (cf., Mowrer and Ullman, 276, for a discussion of these problems.) The whole problem is further complicated by the temporal order of multiple consequences. What happens if we give reward followed by punishment? Punishment followed by reward? Many problems of punishment await empirical investigation at all levels of learning complexity.

3. Several findings have been difficult for goal-reinforcement theories to cope with, e.g., sensory pre-conditioning (Chapter XI), partial reinforcement (Chapter XI), and latent learning, and the next few years may very likely see further experimental work or revision of theory concerning these problems.

4. Pseudo-conditioning (Chapter XI) has forced a revision of methods of calculating amount of true conditioning, but it also has posed an interpretive problem. The best evidence at present

is that it involves generalization, but much more experimental work is needed.

5. In multiple-response serial learning, the nature of remote associations and the explanation of the bowed serial-position curve represent challenging problems. The current activity on spread of effect in serial lists augurs well for a fairly complete analysis of this phenomenon in the near future. Intra-task similarity is in need of much further experimentation, since probably its potency as a determinant of rate of learning has been largely overlooked.

6. It is to be hoped and probably to be expected that there will be a gradual movement of research concentration toward the more complex forms of learning, especially thinking.

7. Since this book is concerned with environmental variables manipulated by *E*, we have concerned ourselves neither with problems relating to individual differences, nor with forms of naturalistic observation wherein *E* observes learning behavior as it occurs naturally during the normal course of events. Very little of the latter type of research has been attempted and it may be, as Hilgard has suggested (151), that we could learn a great deal about learning if we followed a child around for several hours, noting everything he did.

SUMMARY

1. We have given illustrations of the ways in which current learning theories differ in their guesses as to the characteristics of the processes which constitute learning.

2. The discussion was organized around what we called primary variables. These variables are presumed to be critical variables in that without a certain minimal amount of each, no learning will take place. The primary variables were: (1) motivation; (2) consequences; (3) frequency; (4) contiguity. We referred to the variables studied in the previous chapters on learning as secondary variables in that they affect the rate of learning only within limited amounts.

3. Three learning theories, goal-reinforcement, perceptual-learning, and contiguous-conditioning, were evaluated as to their handling of the primary variables.

4. Experimental studies designed to test the validity of the theoretical points of view were sampled. The experiments included: (1) tests of the contiguous-conditioning theory; (2) latent learning; (3) place vs. response learning; (4) spread of effect; (5) the rôle of frequency; and (6) the rôle of contiguity.

5. Finally, we summarized what appear to be the main topics for future critical research.

CHAPTER XV

Forgetting — Retention

INTRODUCTION

It cannot be denied that forgetting is an important practical problem in everyone's life. It is also one of the most irritating of all psychological phenomena because *not* forgetting is esteemed and rewarded so highly in our society. By reason of failure to recall a man's name we commit a social blunder; for our inability to remember who discovered America we lose the esteem of our colleagues as well as 64 dollars; for failure to regurgitate the professor's ideas at examination time we fail a course.

The dual title given this chapter indicates that there are two ways of regarding measurements of the same psychological process. *Retention* is a term referring to the positive aspects of memory, *forgetting*, a term referring to its negative aspects. Retention and forgetting are thus reciprocal terms for the quantitative aspects of memory, with retention indicating the amount *remembered* under specified conditions and forgetting the amount *not remembered* under the same conditions.

Learning and forgetting. The study of conditions which influence forgetting (or retention) logically follows the study of learning. Having required *S* to learn a response to a given degree, or having asked him to repeat a response *N* times, we should concern ourselves next with the conditions which cause the functional strength of that response tendency to decrease. Indices of forgetting are obtained by measuring the strength of response tendencies a given time *after* learning to a criterion. In most cases the measured strength of response will be less when time is allowed to elapse following learning than at the time of reaching the criterion. The amount of forgetting is indicated by the

difference between scores at the end of learning and after the lapse of a specified interval.

Although learning implies "gain" and forgetting "loss," a little reflection will show that consecutive measures of performance increment (from which we infer learning) also indicate what is *not* forgotten from trial to trial. In verbal learning, for example, the number of correct responses given on a trial indicates the amount retained of the learning which took place on the previous trial or trials. The performance measurements from which we infer learning are actually retention measurements. As a convenience we speak of measuring retention and forgetting sometime after learning, but we recognize that we are not thereby isolating psychological processes. This recognition allows us to establish an extremely important predictive principle. Let us examine this principle and the rationale for it.

Predicting forgetting from learning. The process which causes forgetting is not initiated only after learning ceases; it is present during learning but becomes clearly evident only when active learning stops. We may assume that two processes are involved in learning as we record it. One of these processes tends to strengthen response tendencies (learning) and the other to weaken response tendencies (forgetting). When we measure an increment in performance we are measuring the net effect of two simultaneous processes, with learning greater in magnitude than forgetting. The reverse of this may be seen in individual learning curves in which performance decreases temporarily. At least a portion of this decrease may be attributed to the temporary superiority of the forgetting process. Indeed, it may be assumed that without the concurrent forgetting process during acquisition of a response, the rise in the learning curve would be markedly accelerated.

From slow learning we may infer a forgetting process of greater magnitude than we infer from fast learning. Since the forgetting process continues after active learning, we arrive at a prediction of rate of forgetting from an observation of the rate of learning: *when learning is rapid, forgetting will be slow, and when learning is slow, forgetting will be rapid.* This principle, while not universal, holds over a wide range of conditions and if kept in mind will greatly facilitate the study of forgetting.

Variables which influence the rate of learning may be expected to influence also the rate of forgetting; variables which *increase* the rate of learning will *decrease* the rate of forgetting.

Plan of chapter. In studying forgetting we need no introduction to the materials used, since for obvious reasons these are the same as in studying learning. Usually, multiple-response tasks have been used, only a small amount of work having been done on retention of conditioned responses and thinking responses. The three main sections of this chapter are:

1. Methods of response measurement
2. Variables influencing rate of forgetting
3. Theory

METHODS OF RESPONSE MEASUREMENT

The measurement of retention and forgetting is complicated somewhat by the several different methods which have been employed, since each method yields a different quantitative score and more than a single method may be used in one experiment. Let us consider separately the methods currently used.

Recall measures. Recall is the most commonly employed measure of retention. To secure this measure *S* is presented with a stimulus and is asked to call up the response which had previously been attached to it. Usually the strength of the stimulus-response tendency has been developed to a known degree before the retention test is given. The stimulus which is presented to elicit the expected response may be very specific or very general, depending upon the material used. With paired associates the retention test would be given by presenting the series of specific stimuli one at a time and instructing *S* to give the response which had been previously associated with each. A limited time for recalling the responses may or may not be imposed. The stimulus may sometimes be very general as when *S* is given the name of a poem and asked to write down or recite the full poem.

The recall method is sometimes used in studying the reproduction of figures or designs and in this context it is called the method of *reproduction.* *S* is shown a design for a short period of time, and after the design is withdrawn asked to reproduce it as nearly

as possible. Successive reproductions after various intervals provide indications of the changes which may take place in retention of the design.

Relearning measures. After a task has been learned to a given criterion, forgetting may be measured after a lapse of time by determining how much time or how many trials it takes to relearn the task. Comparison of the original time to learn with the time to relearn shows how much was retained from previous learning. This comparison yields what is commonly called a *savings score*. If 20 minutes were required originally to learn a task and 5 minutes to relearn the same task, a saving of 75 per cent results. If it took 20 minutes to relearn, the saving would be 0 per cent, forgetting would be 100 per cent. The smaller the saving the greater is the forgetting and the less the retention. This method dates back to Ebbinghaus. It is still used extensively in memory studies, and often is employed concurrently with recall measures, since experiments may be arranged so that *S* continues after the test of initial recall to re-mastery with no halt in the experimental procedure.

Recognition measures. Suppose *S* has learned a specific group of words. If we place those words among a group of *new* words, how many of those learned will he identify correctly? The number correctly identified defines a recognition measure of retention. As can be seen, this measure is subject to considerable variability depending upon the relation of the learned material to the new material into which it is imbedded. For example, let us take an extreme case. If we required *S* to learn a list of adjectives and then placed the adjectives among a group of nonsense syllables, *S* would probably show very small loss in retention. Obviously the similarity of the test material to the other material is an important variable which determines the recognition score. Because of this, the method is not commonly employed in laboratory experiments except for special problems. It might be used, for example, in studying the effectiveness of advertisements. *S* is allowed to study a certain number of advertisements and then asked to pick them out from a much larger group. This demonstrates the use of the recognition method in a manner which would be very meaningful from the standpoint of the advertiser. Multiple-choice examinations, with which most students are

painfully acquainted, make use of the recognition method, and, as we know, the finer the distinction among the various alternatives, the more difficult it is to choose the correct response.

In using the recognition method correction must be made for chance response. If ten words are learned and then put into a group with ten other words for the recognition test, *S* could be expected to get five correct by chance. In scoring recognition tests, therefore, chance response is usually taken to indicate 100 per cent forgetting.

Reconstruction measures. The method used to derive these measures requires that *S*, confronted by the separable parts of a task presented in jumbled order, replace the parts in the order in which they were at the time of learning. Thus, after *S* learns a list of words to a given criterion, *E* shuffles the order of the words and asks *S* to order them as they were at the time of learning. This method is seldom used in experiments.

As previously mentioned, different methods will almost certainly give different results, so it would appear that we have different "retentions." At this stage in our study we are not likely to fall into the error of asking, "But how much is *really* retained?", or "How much is *really* forgotten?". We recognize that scientific phenomena are a function of the methods by which they are measured, and that to ask one of the above questions is meaningless. No one can properly speak of *the* measure of retention or *the* measure of forgetting. Instead, we have at least four different measures, all of which probably tap the same general process, but tap it in different amounts and ways. Which method will be used depends upon the material being employed, and other factors which are considered separately for each experiment. We need only caution ourselves against falling into the trap of trying to make direct comparisons among retention measures obtained by different methods.

FACTORS INFLUENCING RATE OF FORGETTING

Time Between Learning and Retention Test: 1. Basic Facts

In studying forgetting *E* allows a period of time to intervene between the learning and the retention test. This period is longer than that normally given between trials during learning, and

this procedural distinction (plus instructions to *S*) becomes the basis on which learning and retention are distinguished. In studies of retention, therefore, time becomes a stimulus variable—a condition which *E* manipulates. Retention tests may be given at any time following learning, and we may expect changes in retention as a function of this time interval.

Because of the common observation that, as time passes, more and more is forgotten, it is easy to fall into the error of believing that time *causes* forgetting. Time is directly related to the amount of forgetting in most situations, but we use it only as a low-level explanatory concept. McGeoch (232) clearly demonstrates that passage of time allows for the operation of processes which cause forgetting, but that time in and of itself does nothing and is not to be accepted as a basic causal agent in psychological explanations. Rather, the quest is toward the identification of the processes which take place in time and which *do* cause forgetting. Time has the status of an experimental variable—and an important one—but we will look to other factors when we attempt to explain forgetting.

Ebbinghaus retention curve. The first systematic attempts to measure forgetting as a function of time were made by Ebbinghaus (83). Indeed, the "Ebbinghaus retention curve" is a psychological classic. That the work was sound has been demonstrated by many other *E*s who have confirmed the work in principle. Ebbinghaus, we will remember, served as his own *S* and multiplied the *N* by learning many lists under the same conditions. Retention was measured in terms of relearning (savings) scores, and nonsense syllables were his chief material. The results of his extensive study are summarized in Fig. 83, in which both the retention and the forgetting curves are plotted as a function of time.

The basic generalization concerning the curve of Fig. 83 is that forgetting is most rapid immediately after learning and then proceeds at a slower and slower rate as time passes. Except for certain special conditions (which we shall consider shortly) this basic law has been confirmed many times, although the actual amount of change in retention with time will vary as a function of several variables. By and large, other *E*s using materials comparable to those employed by Ebbinghaus have found that the initial drop in the curve of retention is not as great as that shown

in Fig. 83. These *Es* have used groups of *Ss*. In view of the fact that Ebbinghaus had learned literally hundreds of nonsense syllables during the course of his experiments, it is reasonable to believe that the initial drop in his curve is still further increased by the additional interfering tendencies present in his memorial background.

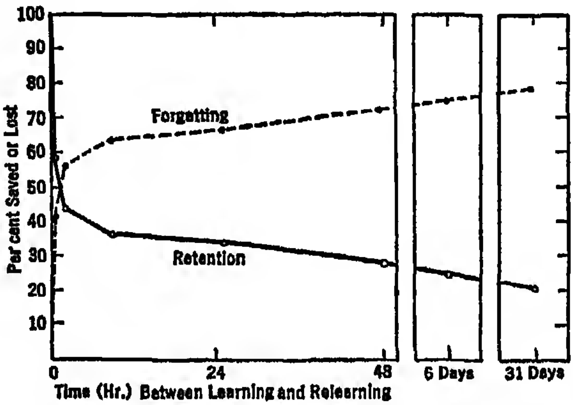


FIG. 83. Forgetting and retention as a function of time. Data from Ebbinghaus (83).

Retention of maze habits by rats. The initial rapid drop in the retention curve is usually found in human maze studies and has been shown also in retention of maze habits by rats. The latter may be illustrated in a study by Bunch (49). Rats were placed in a 14-unit multiple-T water maze. Five groups of rats mastered the problem to a criterion of three successive errorless trials. One group then relearned the maze after 14 days, another after 30, another after 60, another after 90, and the fifth, after 120 days. The per cent saving in relearning the maze is shown for the five groups in Fig. 84. Note that the first retention group relearned after 14 days so that the curve is interpolated between zero and 14 days.

Other relationships. The retention curve for nearly pure motor habits probably drops very slowly with time as compared with the curve for verbal material, although sound data on this matter are not available. We shall discuss later the methodological

problem involved in measuring retention of motor habits as compared with verbal habits.

The forgetting curve of conditioned responses appears to be less sharply accelerated than is the curve for verbal tasks, although in some studies true conditioned responses may have been confused with pseudo-conditioned responses or reflex responses (e.g., the beta eyelid response, 118).

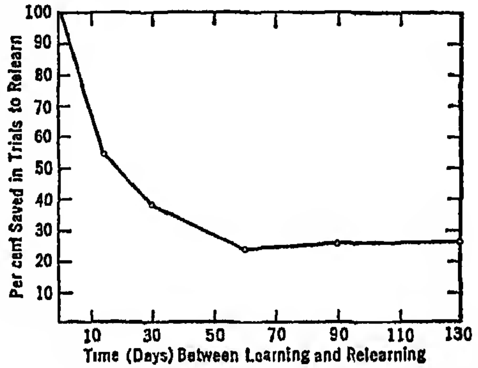


FIG. 84. Retention of a maze habit by rats. Data from Bunch (49).

Let us consider two special phenomena which are associated with the time variable. Both provide special problems of experimental procedure as well as special interpretive problems. The first of these is *reminiscence*, the second is what we shall call *changes in what is remembered.*

Time Between Learning and Retention Test: 2. Reminiscence

Reminiscence is defined as "an improvement in performance, as shown by some measure of ability to recall at some time after the original practice without (any) intervening practice" (56, p. 314). As this definition indicates, reminiscence is the antithesis of forgetting—it is *better* retention rather than *poorer* retention after an interval of time.

Early studies which allegedly discovered reminiscence are open to criticism on the grounds of a flaw in method. In these studies *S* was given a certain number of learning trials, then an immediate retention test, followed later by a second retention test. If the

second test showed greater retention than the first, reminiscence was said to have been demonstrated. Let L1, L2, and so forth, stand for successive learning trials, and let RT1 and RT2 symbolize the two retention tests. The procedure was as follows:

$$L_1 \quad L_2 \quad L_3 \quad L_4 \quad RT_1 \text{———rest———} RT_2$$

If the second retention test produced better performance than the first, reminiscence might at first seem to have been demonstrated. But, note the error involved. The first retention test provided a learning trial. If the second retention test had been given immediately after the first, it might have shown greater retention than occurred after a period of rest (43). If this were the case, reminiscence would not have been demonstrated. It becomes apparent that we need a control condition in order to determine truly whether or not reminiscence occurs. Using the same symbols as above, our two groups should be put through the following procedure:

Experimental Group:	L_1	L_2	L_3	L_4	——rest—— RT
Control Group:	L_1	L_2	L_3	L_4	RT

The difference between the two groups on the single retention test would be the basis for determining whether or not reminiscence took place. If recall for the experimental group is greater than for the control, reminiscence has been demonstrated. The two conditions, of course, need not represent different groups of *S*s since by counterbalancing or randomizing conditions the same *S*s may serve in both conditions.

Because of the procedural flaw in the early studies of reminiscence, the asserted interpretation of the results, at least as they pertain to reminiscence, is in doubt. Reminiscence may or may not have occurred; there is no way of telling. We shall, therefore, limit our discussion to more recent studies which are not marred by this methodological error.

If *S* learns a task to partial mastery and then rests before recall and continued learning (relearning), we have the basic operation which defines the study of the effect of distributed practice (provided, of course, we have a control group which did not receive the rest). It becomes apparent that the study of reminiscence and distributed practice are closely allied, one being a learning

phenomenon (effect on speed of learning) and the other, a retention phenomenon (effect on recall). Because of the near identity of the conditions under which the two phenomena are studied, it has wisely been suggested (56) that we limit the term *reminiscence* to the improvement in *recall* following rest and reserve *relearning trials or trials to reach a given level of mastery following rest* as the basic reference measure for determining the influence of distributed practice. Such a distinction will be used here with the understanding that both phenomena will probably turn out to have a common explanatory basis.

Ward's study with nonsense syllables. The earliest systematic study of verbal learning and reminiscence was published by Ward (399) in 1937, and his study has served as the prototype for later researches. In Ward's extensive experiment, serial lists of 12 nonsense syllables were used as learning material, and two conditions were varied by the use of several groups of *S*s. These two conditions were (1) level of learning before the introduction of the rest period, and (2) length of the rest period. Level of learning was varied in two ways: *S* in one case was given rest after acquiring 7 out of 12 correct responses on a single trial; in another, he rested after having recited the list once perfectly. For each of these two conditions, rest intervals of 30 seconds, 2, 5, 10, and 20 minutes were used. Relearning in all cases was carried to a criterion of 2 successive perfect trials. The basic control condition was the learning of a list to a criterion of 2 perfect trials without any rest period other than the 6 seconds which intervened between trials. Reminiscence is shown by differences in mean recall on control and experimental conditions (Fig. 85).

Ward's recall data for the experimental conditions—as compared with the control—showed that the greatest reminiscence occurred at about a 2-minute interval when *S*s learned to either level of mastery. These results have been supported by Hovland's (e.g., 157) work with nearly identical materials and methods.

It may be wondered how Ward could get reminiscence on conditions in which *S* learned the list to mastery (one perfect trial) before the rest period. Under such conditions what is there to be "reminisced?" The facts are that even though *S* reaches a criterion of one perfect trial, there is a fairly high probability that he will not get the items all correct on the following trial.

Indeed, on the 6-second condition *Ss* got only a mean of 9.75 items correct on the trial immediately following the first one on which all 12 items had been correctly anticipated. Hence, any gain above the 9.75 mean score after longer intervals would be indicative of reminiscence.

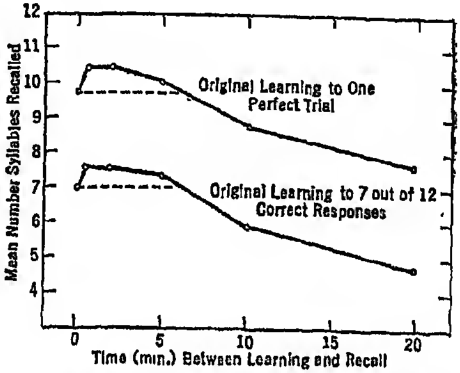


FIG. 85. Reminiscence as a function of time and two degrees of original learning. Area above dotted lines indicates reminiscence. Data from Ward (399).

Reminiscence in pursuit-rotor learning. Reminiscence is a common phenomenon in pursuit-rotor learning. It will occur after varying degrees of learning and after rest intervals lasting from a few minutes to several hours. We shall consider briefly a study by Ammons (4) as representative of methods and results. He used 35 groups of *Ss* (Design Method II), varying both the amount of practice before the rest interval and the length of the rest interval. Take, as an example, one series of conditions on which *Ss* were given 17 minutes *continuous practice*. Following this work period, rest intervals of .33, 2, 5, 10, 20, 60, and 360 minutes were introduced for 7 groups of 14 *Ss* each. The amount of increase in time on target from the 17th to the 18th minute provides an index of the influence of rest intervals introduced between these 2 periods.

The results (Fig. 86) demonstrate that reminiscence increased as the interval increased up to 10 minutes, after which it remained relatively constant for at least an hour. With 6 hours rest there was a decrease in reminiscence, although the amount of

reminiscence after 6 hours rest was still greater than the amount after 20 seconds rest.

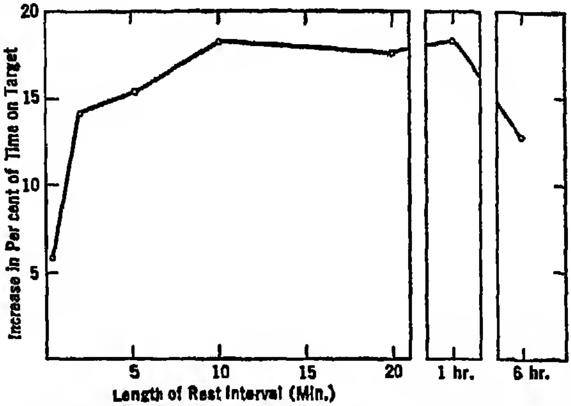


FIG. 86. Reminiscence in pursuit-rotor learning as a function of length of rest interval. Data from Ammons (4).

Ammons' study, when compared with Ward's study, seems to imply that the length of rest interval after which reminiscence will occur in motor learning is much longer than that for verbal learning. The facts as they stand at present with regard to reminiscence in verbal learning are that the phenomenon appears to be dependent upon a highly restricted set of conditions, an analysis to which we shall turn at this point.

Problems of reminiscence. Buxton (56) has called reminiscence a "now-you-see-it-now-you-don't" phenomenon as far as verbal materials are concerned. The reason for this is that some *Es* have been unable to secure evidence of reminiscence under conditions which seem highly favorable for its appearance. The most consistent failures have occurred when serial lists of adjectives were used (57, 256, 415), and the failures occurred despite the fact that other conditions in the experiments were quite comparable to the conditions under which Ward and Hovland, using nonsense syllables, found reminiscence. These facts suggest that for verbal learning a very critical variable is meaningfulness, with reminiscence occurring only with low meaningfulness.

One of the difficulties of flatly stating that meaningfulness is the critical dimension is that no studies have been done in which

meaningfulness was varied and intra-task similarity held constant. Furthermore, as we previously pointed out, it is not a simple task to vary meaningfulness independently of intra-task similarity. The more meaningful the items taken individually, the more probable it is that intra-task similarity will be low. It is possible, however, to have high meaningfulness and different degrees of intra-task similarity if E constructs his materials with that objective in mind. It is very likely that the critical dimension, then, is intra-task similarity and not meaningfulness (57), with reminiscence occurring at high degrees of intra-list similarity. Previous studies may not have found reminiscence with meaningful material because intra-list similarity (unless specifically controlled) is likely to decrease with increasing meaningfulness.

There are several minor problems relating to research on reminiscence, e.g., rest interval activity, spelling vs. pronunciation, but space does not permit us to discuss them. Buxton's (56) recent review contains the best summary of all these problems.

Reminiscence was originally interpreted by the same theory used to explain distributed practice (differential rates of dissipation of positive and negative states) as analyzed in previous chapters. However, the more recent evidence implies that such a theory may be inadequate for both distributed practice and reminiscence and that an adequate theory must relate the concept of generalization (produced by intra-task similarity) to both phenomena (57).

Time Interval Between Learning and Retention Test: 3. Change in What Is Remembered

If *S*s are asked to recall material learned earlier, they may be able to reproduce it perfectly, partially, or not at all. In addition, it is commonly observed that *S* may produce a *given response* in a partially correct way. Partial recall of a response is to be expected for all types of material and in a quantitative sense represents a partial memory error, on a response dimension with perfect recall at one end and zero recall at the other. In verbal material, for example, if the correct response is "aflame," *S* may respond with "flaming"; or, *Underwood* may be recalled as *Woodward*. If a person is asked to recall a prose passage certain essential ideas may be present in his recall, although the precise wording may

vary from the original. If S is presented with a geometrical design for a few seconds and then asked to draw it, his reproduction may be correct in general form but erroneous in detail.

The method of quantifying the results of retention experiments is largely up to E. He may count any deviation from the original response as being "wrong"; he may give credit for partial responses, or he may study the partial responses for what they can tell him about the memory process. We shall, at this point, concern ourselves with the latter procedure.

Two methods may be used to study changes in what is remembered. One method is that of *successive recalls*, in which the same S successively recalls or reproduces the same material after various time intervals. E observes the changes which take place from recall to recall. A second method is used to eliminate the influence of one recall on a subsequent recall by employing a different group of *S*s for each time interval. The changes which take place may be attributed to factors other than those involved in previous recall attempts. The two methods yield different results (139).

Bartlett's work. Some of the most elaborate studies of changes in retention with time have been published by Bartlett (12), a British psychologist. Bartlett has used many different materials, and although he has not insisted upon rigid experimental controls or quantified his data to any extent, he has brought out certain factors which may enter into the observed changes in successive recalls. We shall consider a few of his procedures and results.

In one series of experiments Bartlett required *S*s to read a short passage of fiction. The passage was characterized not only by its disconnectedness (little continuity of events), but also by dramatic content. S read the passage twice and immediately reproduced it as best he could. Then reproduction attempts took place after a day, after several days, and so forth. As would be expected, successive reproductions brought about changes in the content of the story. For one thing, *S*s seemed to strive to make the story have meaning—to avoid the disconnectedness which was apparent in the original. To do this, ideas were imported, i.e., made up, and inserted at appropriate places. Detail seemed to be remembered if it fitted S's original interests and attitudes; if not, the

detail was forgotten or changed so that it coincided with interests and attitudes. Changes, rationalizations, omissions, and so forth, all seemed to take place in accordance with each *S*'s general frame of reference.

From analyzing the successive reproductions in a rather subjective fashion, Bartlett believed that *S*'s personal attitudes and interests caused these changes. However, we need not doubt the *general* finding. As we successively recount our own experiences, we may make them become less and less like the original and more and more a product of our "fill-ins." The fill-ins are no doubt determined by our interests and attitudes toward the experience and by what we wish the recounting of the experience to accomplish. The memory errors of court witnesses are well known. It is customary at present to attribute these alterations and distortions to the individual's frame of reference, about which we shall have more to say later.

In another experiment Bartlett presented *S*s with a series of cards on which were printed signs (figures or drawings) and also a word for which the sign was to stand. The task was essentially a paired-associate learning problem. *S* was told that he was to learn to connect the sign and the word so that if the word were given him he would be able to reproduce the sign. *S* was allowed 7 minutes for study, or less if he felt he knew the pairs before the time was up.

Bartlett took a recall after 15 minutes and then every 2 weeks for some time thereafter. His method of measuring recall was ingenious. He dictated a story for *S* to copy and whenever a word occurred in the dictation which was among those for which a sign had been learned, *S* was supposed to use the sign instead of the word.

Among Bartlett's findings was the fact that signs which had no meaning in and of themselves were among the first forgotten. If *S* seemed unable to relate a symbol to his background of experience, it was readily forgotten. For example, a straight line with two cross-hatches stood for *wind*, and this sign was quickly forgotten. On the other hand, a drawing of an eye stood for *eye* and this was remembered very well. If a word had an already used conventional sign attached to it, then the experimentally learned sign, on successive recalls, came to be more and more

like the familiar conventional one. Dots, used as eyes in representation of a face, became more "eyeish"-looking with successive reproductions. This illustrates the influence of previous learning on the changes in what is remembered. Details which were not clearly attached to the main figure often disappeared unless the detail was particularly novel, in which case it was especially resistant to change with successive reproductions.

Bartlett has shown that rather drastic changes may take place in what is remembered—but because of the non-quantitative nature of Bartlett's results and because of the possible confounding of variables (learning time, intra-task similarity, and so forth), it is difficult to indicate exactly what causes them. Nevertheless, Bartlett's pioneering work abundantly demonstrates that previous learning is the prime factor in effecting the retention changes in other learned material. Here "previous learning" must be broadly interpreted to include also the attitudes, interests, motives, and so forth, which are a part of *S*'s acquired repertoire of responses.

Successive reproduction vs. single reproduction. Hanawalt (139) used geometrical figures, nonsense figures, and so forth. After *S* had studied the material, he recalled it at various intervals of time following the study period. From one group of *S*s immediate recall was required and then successive recalls after 1, 4, and 8 weeks. This is the method used by Bartlett. However, Hanawalt extended his study by using several additional groups of *S*s, asking them to recall only once following the immediate recall. One group thus recalled after 1 week, another after 4 weeks, and another after 8 weeks. These groups had been equated in ability before the major experiment started.

The difference in the results produced by the two procedures —successive reproduction by one group vs. single reproduction by different groups—was quite marked. With successive reproductions by the same *S*s, the frequency of correct reproductions remained almost constant, whereas with corresponding time intervals for different groups of *S*s an Ebbinghaus-type curve resulted. Although Hanawalt found changes in the figures of the successive-reproduction group, he was unable to discern any systematic trend. He believes that the changes observed by other *E*s, e.g., Bartlett, are largely a function of the successive-

reproduction method rather than a function of some intrinsic change within the memory itself. Changes took place, it was true, when Hanawalt's *Ss* made only one delayed recall, but for longer intervals these changes were so considerable that the reproduced figure bore little resemblance to the original, and there was no consistency from *S* to *S* as to the nature of the change.

Another interesting discovery in Hanawalt's study was noted during the learning session when *S* was required to reproduce the figures with the figures themselves directly in front of him. Even under these conditions *S*'s reproductions were not duplicates of the design. Instead, he tended to sharpen certain features and modify others. Such changes, had they gone unnoticed, might be attributed to changes in what is remembered. Hannawalt, however, shows that some of these changes would be present as a function of perceptual inaccuracies and simple inability to draw the figures from a pattern, with no need to invoke changes in memory to account for them.

Very strikingly, these studies and others have shown that changes in what is remembered take place with the passage of time. We have no proof that these changes indicate the operation of mechanisms which are different from those which produce complete forgetting of a response. Forgetting is not an all or none affair; it may be of all degrees. Changes in what is remembered represent an intermediate degree between perfect reproduction and no recall.

Degree of Learning

Other things being equal, the greater the degree of learning the better will be the retention. This is one of the more obvious variables determining the amount of forgetting with passage of time, and we need to give it only enough attention to understand the methods of analysis. Ebbinghaus' study of this variable, some of the results of which are presented in Table 30, are as clear-cut as any more recent data. His degree of learning was determined by the number of repetitions of the lists of syllables, and retention was measured in terms of time saved in relearning. The data show clearly that with the increase in number of repetitions of a list, there is an increase in savings scores after 24 hours.

TABLE 30

INFLUENCE OF DEGREE OF LEARNING ON RETENTION AFTER 24 HOURS

Number of Repetitions	*Relearning Time (Sec.) After 24 Hours*	*Time Saved (Sec.)*
0	1270	—
8	1167	103
16	1078	192
24	975	295
32	863	407
42	697	573
53	585	685
64	454	816

Data from Ebbinghaus (83).

The influence of degree of learning on retention may be determined when the material has been learned past a criteron of one perfect trial. Thus, *S* may be requested to keep repeating the list (overlearning) even though he repeats it perfectly. Krueger has demonstrated that continued repetition of a list beyond the criterion of one perfect trial will enhance both the recall and relearning scores after various time intervals (205).

In dtermining relationship between degree of learning and retention, another method involves working with each stimulus-response connection. This may be illustrated by unpublished data gathered by the author. Lists of paired associates were learned by the usual method in which the order of pairs was varied from trial to trial, so that serial associations were largely avoided. Retention measurements (recall) of the ten pairs of adjectives composing each list were made 20 minutes after original learning. When results were analyzed, the strength of a given response tendency was not determined by the number of repetitions of the list; rather, it was determined by counting the number of times the response had been anticipated correctly during learning. It was assumed that each time a response was correctly anticipated, an increase in the strength of that stimulus-response connection took place attributable to the reinforcing effect of being correct—the O.K. reaction. Hence, all pairs which had been given zero reinforcements (had never been anticipated correctly during learning) were put together; all of those which had been

given one reinforcement were placed together, and so on. To increase the number of cases, those items with 3 and 4 reinforcements were combined as were those with 5, 6, and 7 reinforcements.

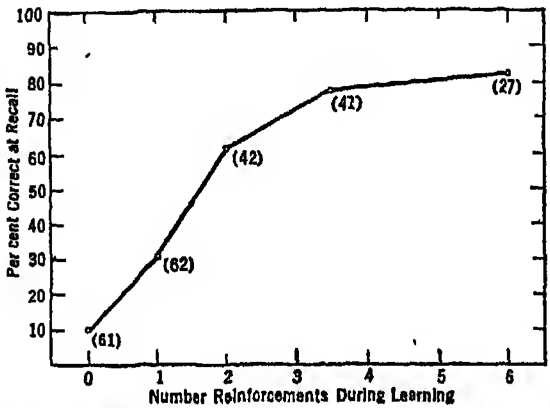


FIG. 87. Retention as a function of degree of learning. The interval between learning and recall was 20 min. The number of cases on which the values are based is shown in parentheses.

The relationship between number of reinforcements and forgetting is shown (Fig. 87) by the per cent correct recall for each degree of reinforcement. With zero objective reinforcements during learning, 11 per cent of the items were correct after 20 minutes, thus indicating that the initial reinforcement process is covert, i.e., reinforcement may take place without *S* overtly anticipating the response. The curve as a whole shows that there is a direct relationship between degree of learning (as defined by number of reinforcements) and amount retained.

Type of Material

In studying the influence of type of material on multiple-response learning (Chapter XII), we referred to three variables: (1) meaningfulness, (2) intra-task similarity, and (3) affective tone of material. In evaluating type of material as it influences retention we shall add a fourth task variable, amount of motor involvement.

Meaningfulness. We have seen that meaningfulness profoundly affects speed of learning. We have said (in the introduction to the present chapter) that speed of learning and speed of forgetting are inversely related—the faster the learning, the slower is the forgetting. According to this principle we should expect that the greater the meaningfulness, the slower is the forgetting. In so far as can be told at present this relationship seems to hold, although no careful study has been reported in which meaningfulness has been varied while holding other variables constant.

In view of the fact that we do not have satisfactory data for several points along a dimension of meaningfulness, we shall cite evidence which leads one to infer the general relationship just mentioned. Reed (303), in the study of concept formation reported in Chapter XIII, compared his results on retention of these concepts with Ebbinghaus' results for nonsense syllables. Reed showed that in a great many respects his procedure—except for the material—was quite similar to the procedure used by Ebbinghaus. After six weeks Reed found only 10 per cent loss in retention of concepts as compared with almost 80 per cent loss in the case of Ebbinghaus' nonsense syllables. The curve of retention of poetry shows a form which is similar to the nonse syllable retention curve, but it never falls as fast or as far over comparable time intervals (406); retention curves of factual material show about the same relationship (80). The retention of substance material (ideas which cannot be derived from a single sentence) shows only slight loss over a period of 80 days (34). There seems to be enough evidence to accept the general principle tentatively, although this evidence is far from systematic, and we cannot be sure that other factors, such as intra-task similarity, have been comparable for the various tasks in these studies briefly cited.

We should discuss some of the difficulties involved in designing a systematic study of meaningfulness as it influences forgetting. Suppose we have two sets of materials which vary only in meaningfulness; intra-task similarity, affective tone, and so forth are the same for both lists. We wish to determine whether there is a difference in retention of the two lists. Obviously, our first step is to require *Ss* to learn the lists. Since the two lists vary in

meaningfulness, we know that they will be learned with different speeds. Consequently, we cannot use a constant number of trials or time for both lists, for the degree of learning attained would be different and degree of learning as such effects rate of forgetting. An alternative would be to have both lists learned to the same criterion, perhaps one perfect trial. This would be fairly satisfactory, but would still not provide an accurate equation of degree of learning. Let us look closely at the reason for this.

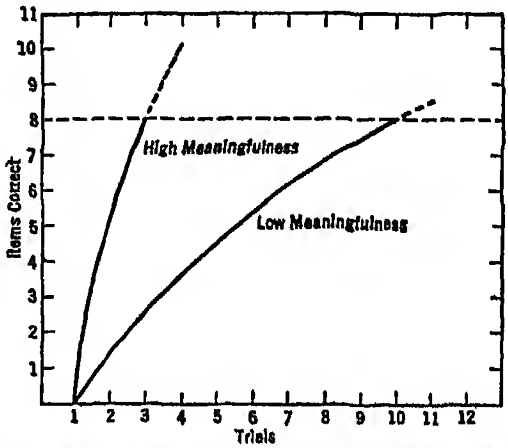


FIG. 88. Hypothetical illustration of difficulty in equating degree of learning of materials learned at different speeds to the same criterion.

If the two lists are learned to the same criterion but at different speeds, the slopes of the two curves are different. An extreme hypothetical case is pictured in Fig. 88. Both lists, one of low meaningfulness and one of high meaningfulness, were taken to a criterion of 8 out of 12 correct responses. Let us say only two *S*s are involved, known to be of equal learning ability. The list of low meaningfulness is learned to the criterion in 3 trials, the list of high meaningfulness in 10 trials. Is the degree of learning equal? If we were to extend the curves for an additional trial, we would in effect be measuring how much each *S* had learned up to and including the criterion trial. When we do this, we see that the degree of learning for the list of low meaningfulness is actually less than that for the list of high meaningfulness. The figure

shows that with a list of high meaningfulness, *S* would get 10 correct on the next trial, whereas with low meaningfulness, he would have only 8.5 correct.

Unless the speed of learning of the two lists is drastically different, as in the illustration, the error resulting by having a mastery criterion is not large, but nevertheless it is an error. We can see that a more satisfactory equalization of degree of original learning could be accomplished by setting the criterion of learning for the list of high meaningfulness slightly less than for the other, so that the extended curves would show approximately equal degrees of learning if a subsequent trial were given. *This problem, the problem of equating degree of learning, is present in all studies of retention where type of material is varied.*

Another difficulty should be mentioned briefly. In long term retention studies with highly meaningful materials, it is possible that *S* may rehearse during the retention interval. There is really no satisfactory way of checking this since *S*s' reports on this matter are always open to question. He may forget not only the material itself but also that he rehearsed. If rehearsal is reported our relationship is confounded. Instructing *S* not to rehearse may be an invitation to rehearse. For careful derivation of a law concerning the relationship between type of material and degree of forgetting, it would be better if the retention tests for different materials could be made after short rest intervals during which *S*s' activity could be controlled more adequately. For long term retention studies it must be assumed that rehearsal, if it takes place, is equal for material with different degrees of meaningfulness. This is not a sound assumption to make, especially if the meaningful material is intrinsically interesting—something which few *S*s report concerning nonsense syllables.

Intra-task similarity. We have indicated that it is possible to maintain a constant degree of meaningfulness while varying intra-task similarity. Since intra-task similarity influences rate of learning, we also recognize that the problem of equating degree of learning is the same as discussed for meaningfulness.

One way of manipulating this variable is to use two lists of different degrees of intra-task similarity. Employing this procedure, Gibson (108) has shown that the greater the degree of intra-list similarity the more rapid is the forgetting. The paired-

associate lists were made up of geometrical figures as stimuli and nonsense syllables as responses (see Chapter IX), with the similarity of the stimuli varied.

Another method of manipulating intra-task similarity is to vary the items within a single list. Buxton and Newman (58) prepared two serial lists. In one list were six nonsense syllables and two nonsense-like figures (relatively meaningless line drawings), and the other, six figures and two syllables. Taken individually, all syllables and figures had approximately the same degree of meaningfulness. Where there were six syllables and two figures, it was presumed that the two figures were more set apart or isolated than were the syllables. The reverse would be true in the case of six figures and two syllables. To learn the lists, *S* would have to differentiate among six items of one kind, but between only two of another kind. Results of the study as a whole show that the isolated items were learned more rapidly and retained better than the six items of a kind.

It is a fair conclusion that, other things being equal, the greater the intra-task similarity the slower is the learning and the more rapid is the forgetting.

Affectivity. Previously (Chapter XII) we have examined some of the obstacles to a clear-cut test of the influence of affective tone of the material on learning. These difficulties, of course, will make an interpretation of forgetting following learning equally ambiguous. In as near final form as possible we wish to point out the major difficulties of interpretation which the procedures and results connected with the isolation of this dimension have encountered.

First we should mention that there is no experimental agreement on whether pleasant, unpleasant, or neutral material is forgotten most rapidly. The reasons for this are largely those which were mentioned in Chapter XII—inadequacies of experimental design. A second point to bear in mind is that there is no agreement on the operation which must be performed to define this dimension. Indeed, it may well be that not one but several dimensions are actually being manipulated. Pleasantness and unpleasantness have commonly been used to designate the extremes of the dimension, but the actual placement of the material on the dimension has been attempted in several ways; and

there is no assurance that the different methods are tapping the same process.

Edwards (85) has called for restatement of the problem of the retention of affective experiences, not in terms of pleasantness and unpleasantness as such, but in terms of whether or not the experience or material learned conforms with the individual's frame of reference. He believes that the contradictory results concerning retention of alleged affective experiences may be due to the fact that pleasant experiences need not necessarily conform to the frame of reference and that unpleasant experiences need not be at odds with the frame of reference. For Edwards the degree of conformity to the frame of reference is important—not whether the experience is pleasant or unpleasant.

Frame of reference in this context is rather loosely defined but is assumed to indicate an individual's values and enduring attitudes on various matters. Thus, a speech with a strong communistic leaning would not fit into the frame of reference of a Republican senator from Iowa. It is Edwards' hypothesis that if experiences conform—do not conflict—with the frame of reference, they will be better retained than if they do not so conform.

1. *Edwards' Experiment.* To test the hypothesis, Edwards performed an experiment (84) in which three groups of 48 *S*s each were singled out (by an attitude scale) with respect to their attitude toward the New Deal (the experiment was performed in 1940). One group was Pro-New Deal, the second was Anti-New Deal, and third was Neutral. To these groups a speech was read which included statements for and against New Deal policies. Immediately after hearing the speech *S*s were required to take a multiple choice (recognition) test covering the content of the speech. The questions were so constructed that half of the correct answers were Anti-New Deal statements and the other half Pro-New Deal statements. The initial test gave a measure of acquisition. Three weeks later *S*s took the test again; hence, retention measures were derived.

The results of both the learning (immediate recognition test) and forgetting (delayed recognition) are shown in Table 31. This table indicates that the group with the Pro-New Deal frame of reference learned facts (presumably from the speech) which were favorable to the New Deal much more readily than they

did those which were opposed to the New Deal. On the other hand, the Anti-New Deal group learned material consonant with their frame of reference more rapidly than they did material which was not consonant.

TABLE 31

RETENTION AS A FUNCTION OF THE CONFORMANCE OR NON-CONFORMANCE OF MATERIAL TO THE FRAME OF REFERENCE

	Mean Correct Responses by Groups		
Correct Anti-Responses	*Pro-New Deal*	*Neutral*	*Anti-New Deal*
Immediate Recognition	9.94	11.77	13.02
Delayed Recognition	7.71	9.48	11.77
Per Cent Loss	22	19	10
Correct Pro-Responses			
Immediate Recognition	15.73	12.83	10.90
Delayed Recognition	14.48	11.50	9.52
Per Cent Loss	8	10	13

Data from Edwards (84).

The retention measurements require special consideration. In terms of the absolute number of units forgotten over the 3-week period, there is little difference as a function of the three different frames of reference. However, since some consideration must be given to differences in the level of learning attained (as shown by the immediate test), the per cent loss over the 3-week interval would be the most meaningful statistic. The trends in the per cent loss will not satisfy rigid tests of statistical significance, but they conform to Edwards' hypothesis. They also conform with the general statement that the slower the learning the faster is the forgetting. Thus, with the variable, "frame of reference," Edwards has shown the basic relationship between learning and forgetting which has been demonstrated with other variables in the more nearly classical laboratory situation.

Whether or not we need the additional concept of frame of reference for this kind of experiment is a moot point. Should we merely interpret the differences in learning as a function of transfer? Learning material which is highly similar to but not in

conflict with already existing ideas and information would produce high positive transfer; learning material which is opposed to those existing ideas and beliefs would result in straightforward negative transfer. Furthermore, the greater the negative transfer between two activities the greater is the retention loss of either after a period of time. One might well interpret the results of this study in such a fashion. The use of the term, frame of reference, is intended to be helpful in describing the relationship between already existing attitudes or beliefs and the material on which *S* is to be tested. We are, in effect, using the concept to state the similarity relationships, and it is obvious that is not easy to break down such complex things as political beliefs into a unit analysis of the *A-B*, *A-C* type, as we have done in treating transfer of training.

Again, in speaking of these sorts of studies, we ought to point out that quantitative laws of retention are difficult to evolve unless degree of learning before the retention interval is equated. The major finding of the Edwards-type study does not concern retention but learning. Clearly, wide differences in learning take place as a function of the frame of reference or as a function of transfer of training.

2. *Other Approaches.* There are several other kinds of studies directly or indirectly concerned with affectivity. In an effort to produce different rates of forgetting for different materials, Wallen (397) has introduced a technique designed to personalize material. His method, in brief, was to present a series of 40 adjectives to *S*, who judged whether he himself did or did not possess the trait indicated by the adjective. Some time later the adjectives were presented again to *S*, this time with a notation after each adjective. *S* was told that a group of his friends had rated him on these traits and the notations indicated whether they thought he did or did not possess the traits. Actually these supposed friends' ratings were bogus and were derived by *E*'s systematically reversing one half of *S*'s own ratings. *S*, however, fully believed they were friends' ratings. Later *E* asked *S* to recall the friends' ratings. Would *S*'s recall be better of those friends' ratings which were favorable to him, or of those less favorable? Would he recall ratings which conflicted with his own picture of himself or would he "repress" these?

The results of this experiment suggest that there will be greater forgetting of items which conflict with *S*'s own evaluations of himself, than of those which do not conflict. However, the case for a form of selective forgetting based on personal conflict is not established. The differential forgetting which occurs is what one would expect on the basis of negative transfer of incompatible responses, since in effect, *S* had learned two responses to the same stimulus. Since no control for non-personalized interference was used, we cannot tell whether or not something is added to the forgetting beyond that expected from straightforward interference.

Williams (413) attempted to produce different amounts of forgetting by frustrating *S*s. They learned a list of paired associates, half of which were words of an aggressive type (e.g., *fight*, *hate*, *curse*) and half neutral and not related to aggression. Twenty-four hours after learning, *S*s were subjected to a new situation which was so managed that they inevitably failed. As "punishment" for failing they were severely criticized and given an electric shock. Immediately after this experience they were tested for their recall of the list of paired associates. A control group followed exactly the same procedure except they did not experience failure and punishment. Comparing the two groups there was no evidence of selective retention. The aggressive words were recalled no better by the experimental than by the control group.

Sharp (349) used a group of words which were taken from case histories of neurotic patients. One paired-associate list was known to be acceptable to *S*s (e.g., winning-respect), another unacceptable (e.g., going-insane), and another neutral (e.g., flying-kites). Neurotic as well as normal *S*s learned and recalled these lists. Both groups' results show that the unacceptable material was forgotten more rapidly than the acceptable and the neutral lists. However, other *E*s (339) have been unable to confirm these results, and this may mean that a peculiar combination of unequal groups and lists differing in difficulty could account for Sharp's results.

Space does not permit inclusion of other studies. Those briefly reviewed above give us samples of various approaches to the problems of affectivity and related variables. Taken as a whole,

the studies provide very meager evidence of differential forgetting as a function of affectivity. If the procedures outlined in Chapter XII are followed, and if learning is carried to a mastery criterion before the interval, very likely this variable, at least as defined by Edwards, *is* an effective variable. In some cases it is assumed falsely that material will show differences in rate of forgetting without showing differences in rate of learning. The assumption is untenable since it implies a characteristic of the material which "turns its influence on" *after* learning even though the presence of the characteristic was established before learning. We may, of course, introduce conditions after learning which will influence retention, and we might even change the affective characteristics of the material after learning, but unless this is done a characteristic of material which influences retention will also influence learning.

Motor-verbal dimension. This dimension, which can be used to distinguish different types of tasks, is an extremely difficult one with which to experiment. Although we may clearly set up tasks with different amounts of motor and verbal involvement, it is not easy to get comparable units of measurement. Yet this is a necessary step in order to equate the degree of learning of the tasks before making retention comparisons after different time intervals. Let us look at this predicament.

1. *Maze vs. Nonsense Syllables.* McGeoch and Melton (234) have compared maze retention with retention of nonsense lists, one week after learning. Degree of original learning was the same for both the maze and nonsense list (one perfect trial), and approximately the same number of trials was required to reach this criterion on both tasks. Three mazes and three lists of syllables were used, both sets of materials being of graduated difficulty.

These investigators found no consistent evidence that the retention of either task was superior to the other. Further, they suggest that other *E*s have supposedly found a difference in retention between nonsense syllables and maze habits because the degree of learning was not equalized.

The McGeoch-Melton experiment gives a far from satisfactory answer to the question concerning effects of differences in motor-verbal involvement on retention. In the first place, of course,

only two points on the dimension were actually explored. Secondly, it can be argued that these two points are not very far apart on the dimension. The basic process of maze learning is highly symbolic or verbal, since many *S*s learn mazes by memorizing left-right sequences. The motor component may be almost exclusively auxiliary in nature for many *S*s; the arm and finger movements may merely indicate a response determined symbolically in much the same fashion that movements of the larynx are used to indicate a verbal response in learning nonsense syllables. If this be true, and undoubtedly it is with many *S*s, we should not expect a large difference between maze retention and nonsense-syllable retention. The situation calls for precise analysis of how integral a part are the muscular movements in the maze learning process. The Mc-Geoch-Melton experiment has shown that retention of a maze habit does not differ greatly from retention of a list of nonsense syllables. If, however, the maze and the nonsense syllables do not represent significantly different points on the motor-verbal dimension, we do not have an adequate statement of the relationship for which the experiment sought.

2. *Pursuit Rotor vs. Nonsense Syllables.* A recent study (211) was performed in an attempt to get a clear-cut test of whether or not the verbal-motor dimension was effective in producing differences in retention. Pursuit-rotor learning was selected as a task representing almost complete motor involvement and having but little verbal or symbolic involvement. The inevitable nonsense syllables were used to represent the opposed end of the dimension. Both tasks were repeated for 10 trials each, with 30-second rest between trials. A trial for the nonsense syllables consisted of one complete presentation of the 15-unit list. For the pursuit rotor, 30 seconds practice in keeping the stylus on the target constituted a trial. Since the 15 nonsense syllables were presented at a 2-second rate, the total time as well as the number of trials practiced was constant for both tasks. Different groups were used to obtain retention measurements of both tasks after 1, 7, 28, and 70 days.

At the end of the 10 practice trials performance on both tasks was increasing almost linearly. On the 10th trial a median of 5.8 syllables was correctly anticipated and the stylus was kept

on the target an average of 4.7 seconds. Retention measurements showed that in terms of savings scores the pursuit task was significantly better retained than the syllables. The pursuit-rotor scores showed considerable reminiscence, even with allowances for the test-retest error. No such finding was in evidence for the retention of the nonsense syllables.

Unfortunately there are several factors which prohibit a clear-cut interpretation of these results. The groups were poorly matched on original learning scores. For example, on the nonsense list, the group which recalled after 70 days had a median of 4.0 syllables correct at the end of the 10th learning trial, whereas the group which recalled after one day had a median of 6.0 correct at the same point in learning. The one-day pursuit rotor group had a median time of 3.34 seconds on the target at the 10th trial whereas the 70-day group had a median of 8.31 seconds on the target on the 10th trial. Retention comparisons made with such widely divergent degrees of learning are questionable. How to get around this difficulty in the design is an embarrassing question, since, if the groups were matched on one task, they may not necessarily be matched on the other.

A still more puzzling feature of the procedure in this experiment is that of knowing what to make of the measuring units employed to indicate the degree of original learning of the two tasks. If one used the same number of trials did that mean that the two tasks were learned to an equal degree? Theoretically a perfect score on the pursuit rotor would be 30 seconds on the target; a perfect score on the nonsense list would be the correct anticipation of all 15 items on one trial; for a given trial neither task could be performed any better. We know that *S* would eventually learn to anticipate all 15 items correctly—the learning curve would level off. But where does the pursuit rotor learning curve level off? At 30 seconds on target? Even if the *Ss*' performance could be carried to perfection on the pursuit rotor, would that be equivalent to the degree of learning of nonsense syllables which are all correctly anticipated? These questions are not handled adequately by this study; therefore, we must reject any final conclusion that pursuit rotor learning is forgotten less rapidly than is nonsense syllable learning.

It is a real problem to get equivalent degrees of learning for

tasks which are quite diverse in nature and for which the response-measurement units have no common basis for comparison. There is, possibly, one method which might give a fairly satisfactory basis for comparison. Suppose both tasks were carried to a point where, with additional practice, little change in the level of performance resulted, i.e., the learning curves became parallel with the abscissa. For nonsense syllables this would probably amount to several successive perfect trials. For pursuit rotor learning it would require empirical determination; perhaps the average *S* will show little or no improvement past, let us say, 20 seconds on the target out of a possible 30 seconds. Having determined the mean number of trials required for the learning curves of both tasks to level off under given conditions, we run a second experiment. With new *S*s (from the same population as that for which the leveling-off points were determined) we equate the degree of original learning for the two tasks by taking *S*s to a fraction of total learning possible, this fraction being the same for both tasks. Thus, if the optimum mean level of learning attained on the pursuit rotor under given conditions was 20 seconds on target out of a possible 30 seconds, we might use the number of trials to reach 10 seconds on target as a measure of 50 per cent learning. Likewise, when the nonsense syllable curve levels off—slightly under consistent perfection—we could use the number of trials to attain half that perfection as indicating an equivalent degree of learning to the 50 per cent learning on the prusuit rotor. Such a procedure would have to be used on tasks which give the same general shape of learning curve and even then, the comparisons would be rough. It would, however, provide a more precise equation of the degree of learning than is usually attained.

3. *Conclusion.* We must adjudge that we do not have evidence as to whether or not degree of motor involvement is an effective variable in and of itself. Inferential experimental evidence might lead us to conclude that motor skills are retained much better than verbal skills. Pursuit rotor learning, for example, shows little loss (indeed, there is evidence for reminiscence) over several hours. It has often been said that we never forget how to ice skate or how to ride a bicycle. However, we have seen that with highly meaningful verbal material little forgetting is evidenced over

several weeks. Yet, nonsense syllables, even if overlearned, will show rapid forgetting in a short period of time.*

These considerations seem to imply that merely evaluating the verbal-motor dimension is not sufficient; we must also consider the degree of meaningfulness at the same time. Nonsense syllables and substance materials (ideas or facts) may be largely the same with regard to their verbal or symbolic character but vary markedly as regards meaningfulness. Thus, it may be that only when we get low meaningfulness in verbal tasks do we find a more rapid rate of forgetting than we do in motor tasks. Could we manage to equate the degree of meaningfulness for verbal and motor tasks? The answer is probably no at present, and it may be that when we ask the question, "Are motor or are verbal tasks forgotten more rapidly?", we are asking a question which cannot be answered experimentally at our present stage of knowledge and experimental technique. What we can do is to use specific tasks, equate the degree of original learning, measure retention, state the results of the experiment, and decline to go further. Thus, we may discover relative forgetting rates for specific tasks, but until we know the relationship between those tasks better than we do now, we are not justified in generalizing with reference to verbal-motor differences. This dimension remains a crude descriptive one without the status of a stimulus variable for which we can state the response relationship.

Massed vs. Distributed Learning

Two groups learn the same material to the same criterion of learning, but one group learns by massed practice and another by distributed practice. Will retention vary as a function of these two treatments given during learning? The experimental procedure is straightforward so we need follow through on only one study to demonstrate the method and the relationship.

Hovland (160) arranged to have 32 *Ss* serve in eight experimental conditions during which they learned serial lists of 12 nonsense syllables. In four conditions the lists were learned by

* Another line of evidence supports the position that motor tasks are forgotten less rapidly than are verbal materials. The evidence springs from data (54) which show that it is difficult to produce interference between motor tasks. Retention of verbal materials, on the other hand, can be markedly depressed by requiring *S* to learn other conflicting material.

massed practice, and retention was measured after 6 seconds (the usual inter-trial "rest" interval), 2 minutes, 10 minutes, and 24 hours. In the other four conditions learning took place by distributed practice, as defined by a 2-minute rest interval after each learning trial. The retention intervals following distributed learning were, of course, the same as those following massed learning. Learning in all cases was carried to one perfect recitation.

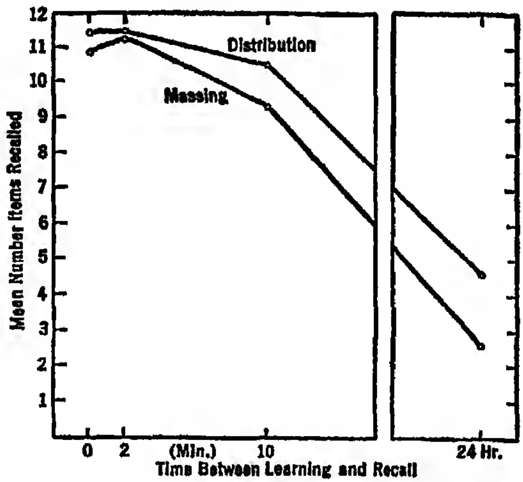


FIG. 89. Relationship between massing and distributed practice and retention. Data from Hovland (160).

Thus, as the design was set up, each *S* learned four lists under massed conditions and recalled and relearned after the four time intervals, and each *S* learned four lists under distributed conditions and recalled and relearned after the four intervals (Design Method IV).

As would be expected, learning was more rapid under distributed conditions than under massed conditions. On the average, about 14.5 trials were required to learn to one perfect recitation under the massed conditions whereas only 9.5 trials were required under the distributed conditions. Previously, we have spoken of the fact that materials learned to the same criterion but at different rates were not perfectly equated. Therefore, a slight error in equalizing the degree of learning probably

existed in this experiment. The error would tend to make the degree of distributed practice learning a little higher than that of massed learning. However, the differences in recall, at least for the longer time intervals, are great enough so that we cannot attribute them entirely to differences in degree of original learning.

The recall measurements are shown in Fig. 89. The difference between massed and distributed practice is significant at all points except for the 2-minute interval, in which the rise in the curve following massed practice is interpreted as reminiscence. A considerable difference appears at the 24-hour recall, with indications that the curves are gradually separating. This trend is more apparent in the relearning scores which are not shown here. We may conclude that even though massed learning requires a greater number of trials to reach a given criterion, the forgetting will be greater in the long run than will the forgetting of the same material learned under distributed conditions. Since similar results have been found with several different kinds of material (233) the generalization is an important one.

Context

All stimuli appear in a context. That context may be thought of, according to one's purpose, as consisting of many components or patterns or of few components and patterns. A stimulus word on a memory drum is printed on white tape background, but this stimulus component is in a larger setting of the memory drum, the room, the sound of the motor, the light from overhead, and so forth. Responses are learned in the usual laboratory procedure with as constant a background context as possible, unless that context is the dimension being explored.

Common observation will suggest that this is an effective variable in producing changes in retention, with the greatest change in context between learning and the retention test producing the greater decrement in retention. As many of you know, the armed forces have a quaint custom which requires one to salute a person of superior rank. Yet an admiral is not saluted while he is in the barber's chair nor a general while he is in swimming. The context had changed to such an extent that the customarily correct response is no longer elicited.

In a sense this variable, change of context, is an "unfair" one to bring up since it does not seem to determine "true" retention; rather, it shows that adding to or subtracting from a given stimulus context to which a response has been learned will reduce the probabilities of that response occurring. Yet, the very fact that we know that stimulus conditions never remain identical, and indeed may vary drastically in real life situations, makes this a very practical sort of experimental variable. Recognition of the potency of such a variable will also serve as a reminder that the three little letters constituting a nonsense syllable are not *the* stimulus but rather the focal point of the stimulus complex. Components of this stimulus complex which remain relatively constant throughout learning will perhaps become associated with the response, and their variations at time of recall may be expected to reduce the frequency of correct response.

It is difficult to vary large portions of the context systematically, and know quantitatively what is being varied. Learning in one room and recalling in quite a different room will bring about a decrement in retention, but it is hard to plot the change in recall as a function of a continuous variable. Hence, laboratory studies have inclined to be more analytical by varying just a small portion of the stimulus context. We will illustrate this with one study performed by Dulsky (82).

Dulsky used paired nonsense lists, ten pairs to a list. Background context was varied by putting the words on different colored papers and then changing the colors at the time of recall. Many different colors were used during learning and many changes were made at recall. As control it was necessary, of course, to have some words retain the same context at recall that they had during learning. We may combine all the conditions which did have a change in context at recall and compare them with all conditions under which there was no change. Under conditions of no change a total of 365 responses were correct at recall, whereas under conditions of change only 265 responses were correct. More startling, however, is the fact that with no change the lists were relearned after 5 minutes in an average of 4.8 trials, but with change it took an average of 13.7 trials to relearn—almost three times as many trials with change of context as without

change. The possibility of interpreting such results in terms of negative transfer is apparent.

Change of context, in short, is an extremely effective variable. The greater the change the greater is the decrement in recall.

Other Learning

Most theories hold that the interference of other learning is a major factor bringing about forgetting; a task is forgotten because other learning interferes with its being remembered. Such theories have as their empirical basis many experiments in which two learning tasks are given *S* in order to see what influence the learning of one (either one) has on the retention of the other.

Retroactive and proactive inhibition. There are two classes of operations by which the effect of other learning or retention is studied. One operation produces *retroactive inhibition,* the other *proactive inhibition.* These may be designated as *RI* and *PI* respectively. We define *RI* as the decrement in retention of a task as a consequence of other learning coming *between* the learning and the retention test. *PI* is defined as the decrement in retention of a task as a consequence of other learning coming *prior* to the learning of the task on which the retention test is made.

In most cases, as we have seen, some forgetting will occur with the passage of time, without the formal learning of other material. Consequently, in *RI* and *PI* part of the decrement in retention is allotted to the other learning under control of *E* and part to "natural" forgetting which would take place without the *formal* learning of another task. Because of this, it is customary to use a control condition in which a single task is learned and recalled. The amount of the difference between natural forgetting and forgetting when another task *has* been learned is a function of the other material. Let *A* stand for one learning task, *B* for another, and dotted lines for time, and our control and experimental conditions are as follows:

	Learn	*Learn*	*Learn*	*Retention Test*
Control (Both RI and PI)	....	*A*	....	*A*
Experimental: RI	....	*A*	*B*	*A*
Experimental: PI	..... *B*	*A*	....	*A*

A is recalled after the same time interval in all three conditions. The comparison of retention of *A* under *RI* condition with the retention of *A* under the control condition gives the amount of forgetting which is due to the learning of *B*. A comparison of the retention of *A* in the control as compared with the retention of *A* in the *PI* condition gives the amount of forgetting that can be attributed to the prior learning of *B*. Comparing the retention of *A* in the *RI* and *PI* conditions would give an indication of the difference in forgetting caused by the locus of *B*, i.e., whether it comes before (prior learning) or after (interpolated learning) the material to be tested for retention. Design Methods II, III, and IV have all been used in studying these phenomena.

It will be noted that the *PI* experimental condition calls for a retention test of the material which was learned second (following *B*). This requires that a rest interval be inserted between the learning and retention test of the second task; otherwise, there would be continuous learning of this second task and no inhibition would be expected. The length of the rest interval must therefore become a rather important variable in the production of *PI*. *RI*, on the other hand, can be obtained by giving *A-B-A* with no appreciable interval between tasks.

We have already seen that comparing the retention under a control condition with the retention under an experimental condition gives a basic measure of *PI* and *RI*. The retention measurements may be any of the four discussed early in the chapter, although in actual practice recall measures have been most widely used. The decrement in retention as a function of other learning may, however, be calculated in several ways. (1) It may be expressed in absolute units. For example, suppose the mean recall under the control condition was six units and under the *RI* condition, three units. The difference, three, indicates the absolute amount of *RI*. (2) The amount of inhibition may also be expressed in percentage terms. To calculate percentage decrement we subtract the retention under the experimental condition from the retention under the control condition, divide by the retention under the control condition and multiply by the usual 100 in order to change to per cent. This formula is: $\frac{C - E}{C}(100)$. In the illustrative problem it would be $\frac{6 - 3}{6}(100)$,

which is 50 per cent *RI*. (3) In some studies the absolute amount of *RI* has not been used; rather, the total forgetting is determined by subtracting the retention score on the *RI* condition from the score made on the last learning trial of that condition. This difference is compared directly with the corresponding difference of the control condition. (4) Latency of response is sometimes used as an auxiliary measure of inhibition. It is assumed that the longer the latency of the response at recall the greater is the inhibition. With human *S*s such a measure is not to be considered a substitute for actual recall measures.

It has often been noted that significant amounts of *RI* and *PI* will be found in recall scores, but that following recall there will be no significant difference in trials to relearn a control and an experimental task. Both *RI* and *PI* are relatively transitory (as in the case of straightforward negative transfer), but if the amount of inhibition at recall is severe, differences in relearning trials will appear.

Important variables. Study of the operations which define *RI* and *PI* immediately suggest the variables which are likely to be important. It is apparent at once that we are dealing with a transfer situation and that factors which influence transfer will influence *RI* and *PI*. Indeed, similarity and relative degrees of learning of the two tasks are the two most important variables determining the extent of *RI* and *PI*. Others which suggest themselves are the length of time interval between various segments of learning, the time between learning and the retention test, the number of tasks which make up *A* and *B*, and so forth.

There will be no attempt here to catalogue or illustrate all variables which influence *RI* and *PI*. Scores of experiments have been performed on *RI*, but only a small number have been done on *PI*, since it is a relatively recent discovery (1927). From the many possible factors which have been shown to influence *RI*, we shall choose three for brief discussion, these three being especially powerful in producing changes in amount of forgetting.

1. *Similarity*. Methods of manipulating similarity are given in Chapter IX. As in the transfer of training situation, unless a minimum amount of similarity is present between tasks, little *RI* will be produced. If conditions of similarity are such that they

produce high positive transfer, they are likely to produce retroactive facilitation also; if similarity conditions will produce high negative transfer, they will also produce high *RI*. The most definitive study on this variable has been performed by Haagen (133), using paired-adjectives. He has shown that with very high degrees of similarity between lists, facilitation will take place in recall of the first list. As similarity decreases, facilitation

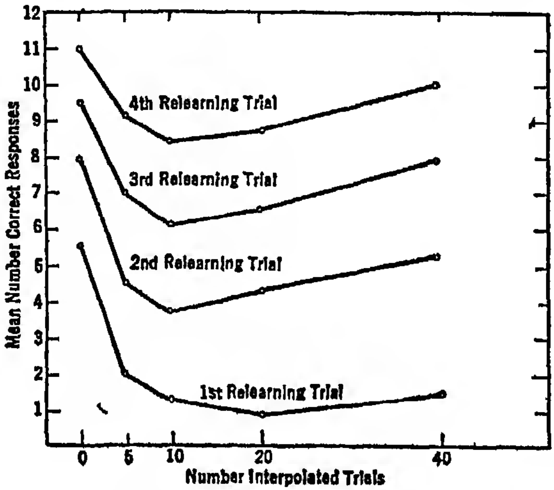


FIG. 90. Retroactive inhibition as a function of degree of interpolated learning. Data from Melton and Irwin (255).

changes to inhibition and the increase in *RI* is very sharp until greatest inhibition occurs when the two lists are of the *A-B*, *A-C* relationship. With dissimilarity of both stimuli and response, only slight *RI* is produced.

No systematic studies on *PI* as a function of similarity have been performed, but one can infer from several studies that it is a very critical variable for *PI* as it is for *RI*.

2. *Degree of Interpolated Learning*. With a wide range of materials, data on this variable are consistent in showing that as the degree of interpolated learning increases, while holding the first list, (original) learning constant, *RI* increases. There is slight evidence that with great overlearning of the interpolated list, *RI* may decrease slightly. Let us look at the results of one study.

Melton and Irwin (225) used serial lists of 18 nonsense syllables. The original list was always presented for 5 trials, and the interpolated list for 0 (control), 5, 10, 20, and 40 trials in five conditions by Design Method IV. Recall and relearning of the original list took place after 30 minutes for all conditions. The mean correct items on the first four relearning trials are shown in Fig. 90. The first relearning trial is, of course, the recall trial. These curves do not show *RI* directly. The amount of *RI* would

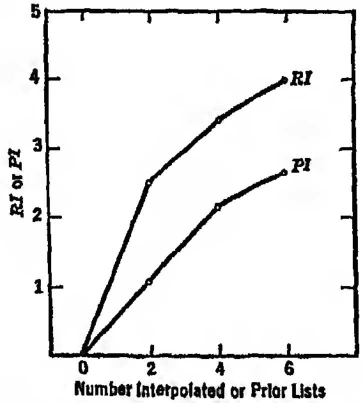


Fig. 91. *RI* as a function of number of interpolated lists, and *PI* as a function of number of prior lists (391).

be calculated by subtracting the mean number recalled under each experimental condition from the number recalled under the control condition. If this is done it will be seen that the amount of *RI* increases with increasing number of interpolated trials up to 20 trials. The amount of *RI* with 40 interpolated trials is slightly less than with 20.

For *PI* the degree of prior learning corresponds to degree of interpolated learning in *RI*. It has been shown that as degree of prior learning increases, *PI* increases (394). The absolute decrement, however, is somewhat less than in *RI* produced by varying degree of interpolated learning.

3. *Amount of Interpolated Learning.* While increasing degree of learning on a single interpolated list produces a marked increase in forgetting, even greater amounts of *RI* results from varying the number of interpolated lists (amount of interpolated learning).

In one experiment (391) the original learning was carried to a criterion of 6 out of 10 responses correct on a list of paired two-syllable adjectives. On four conditions *Ss* were presented 0, 2, 4, or 6 interpolated lists for 4 trials each. Each list had the same stimuli as the original list but different responses. The conditions were counter balanced by Design Method III. The original list was recalled after 25 minutes. *RI*, as shown in Fig. 91, was determined by subtracting the mean number items recalled on the experimental conditions from the mean number recalled on the control condition (zero interpolated lists).

In the same experiment the number of prior lists was varied in exactly the same fashion as were the interpolated lists. The recall results, indicating *PI*, are also shown in Fig. 91. We see that the general relationship between *RI* and *PI* and number of prior or interpolated lists is the same, although the absolute amount of *RI* is greater than the amount of *PI*.

This brief analysis of three major variables shows how large amounts of forgetting may be produced over very short intervals by manipulating other learning. The fact that all these retention losses are underscored by similarity of materials makes interference among activities a major factor in forgetting.

RI vs. PI. All evidence points to the fact that forgetting produced by *RI* is greater than that produced by *PI* over short intervals of time, perhaps, a few hours. However, as the retention interval increases *PI* becomes greater and greater until it equals *RI*. This interval, necessary for the two to become equal, will probably vary as a function of a number of conditions. In one experiment (392), it was about 48 hours.

The high degree of relationship between *RI* and *PI* as a function of manipulable variables strongly suggests that both are based on the same mechanisms and together can account for a great deal of forgetting.

The memory span. The memory span is defined as the number of units which on the average can be reproduced correctly after a single hearing. It has been customary to use one-syllable units of some kind (such as digits) as the material for determining the memory span. *E* reads several series of digits at the rate of one digit per second, and *S* attempts to write each series immediately after hearing it. *S* must reproduce perfectly to receive any credit.

The series read by *E* will vary in length from a short series which all *S*s get correct (perhaps 4 or 5 items) to a long series which few get correct (12 to 15). Each length series is presented several times. If we want a single measure to indicate the memory span for a given *S*, we may (following psychophysical procedure) find the length of series which is correctly reproduced 50 per cent of the time.

RI and *PI* have been discussed in relation to two tasks or two lists. Yet, these same phenomena occur *within* a series such as in the memory span series. Items in a list interfere with the recall of other items in the list, and the interference is a function of the same variables which produce inter-list interference. It is clear that the interference within a list should be a two-way phenomenon; items which come after item *D* should interfere with its recall (*RI*), and those which come before item *D* should interfere with its recall (*PI*).

Summary of Variables

We have discussed all the major variables influencing forgetting as they have been studied by experimental technique. In view of the length of this section a summary of the evidence concerning these variables is in order.

1. *Time.* The greater the length of time between learning and the retention test, the greater is the forgetting.

2. *Degree of Learning.* The less the degree of learning the faster is the forgetting.

3. *Massing vs. Distribution.* If distribution of practice enhances learning (as compared with massed learning), the rate of forgetting will be greater following massed learning than it will following distributed learning.

4. *Type of material.*

   a. *Meaningfulness.* The less meaningful the material, the faster is the rate of forgetting.
   b. *Intra-task similarity.* The greater the degree of intra-task similarity, the faster is the forgetting.
   c. *Affectivity.* The relationship is unknown.
   d. *Degree of Motor Involvement.* Experimentally untested.

5. *Context.* The greater the change in context between learning and retention, the greater is the forgetting.

6. *Other Learning.* The learning of other tasks, especially similar tasks, markedly increases amount of forgetting.

THEORY

A theory of forgetting must start with the basic facts we have summarized and try to account for them with as few assumptions as possible. While no attempt will be made to outline a complete theory here, we may at least look at the general form which the most influential current theory takes.

It is almost axiomatic that because of the close relationship between learning and forgetting, a theory of forgetting must start with at least general assumptions concerning the learning process. In greater detail, the theory needs to explain the relationship between manipulable variables and rate of learning. Since these variables also influence rate of forgetting, the outline of a forgetting theory is automatically established. Only one additional assumption must be made, namely, an assumption concerning "what goes on" between the end of learning and the retention test which may account for the observed decrement called forgetting.

The researches on *RI* and *PI*, on changes in what is remembered, on intra-task similarity, on context, and on meaningfulness, indicate that a key assumption in a theory of forgetting will be one which accounts for interference between activities. That an interaction between activities takes place, and that this interaction produces interference, seems inescapable. Such interference is not limited to serial lists or paired associates. Retroactive inhibition, for example, is not restricted in scope, since it has been produced with many tasks and many organisms. A recent experiment (263) has produced carefully measured retroactive inhibition in cockroaches. Although such a study might be performed partly because those particular *Ss* are plentiful, *E*'s real motive is a desire to test the generality of a phenomenon.

The evidence for interference is by no means entirely inferential. During retention tests *E* observes errors (intrusions) which can be traced directly to other learning. He assumes that such intrusions are indicative of interference. A theory must then proceed to specify what mechanisms may cause the amount of interference

to vary and how, in turn, the mechanisms change with manipulations of the stimulus variables.

The most acceptable theory at present, that of Gibson (106), makes use of *generalization* as the mechanism which produces the interference. A great many data, from both transfer of training experiments and forgetting experiments, support the usefulness of the concept. We have seen that the mechanism is helpful in explaining the results of many transfer of training experiments; we have also seen that retroactive and proactive inhibition are merely elaborations of the transfer of training paradigm.

No one, of course, attempts to explain retroactive and proactive inhibition on the basis of one set of principles, and natural forgetting on another. Retroactive and proactive inhibition are viewed as ways of heightening or exaggerating the processes which are believed to operate in natural forgetting. When *S* learns a single list of words and recalls and relearns that list after one week, forgetting is measured and the assumption is that retroactive and proactive inhibition produced the loss. If we knew all the past learning of *S* (which produces proactive inhibition) and all the learning between the acquisition of the list and its retention (which produces retroactive inhibition), the presumption is that we could account completely for the measured forgetting of the list.

We may, then, use generalization, and its opposition concept, differentiation, to explain forgetting. Differentiation is to be thought of as an integral part of the learning process. Differentiation is the positive process which reduces the generalizing (interfering) tendencies. *The basic assumption of the theory is that following learning* (when active differentiation is no longer occurring) *the generalizing tendencies previously present recover with the passage of time.* On the basis of this assumption one may explain forgetting at three different levels of experimental operation.

1. *Within a task.* Data on intra-task similarity and learning and forgetting support the use of the concepts of generalization and differentiation. During the learning of a task in which components are similar, the components must be differentiated—generalization among the components must be reduced. If the assumption is that generalization tendencies recover with the

passage of time following active learning, these tendencies would produce interference among response tendencies, hence, forgetting.

2. *Between known tasks.* If *S* learns two tasks which are similar (as in retroactive and proactive inhibition), the first task generalizes with the second. Differentiation is established as the second task is learned. With the passage of time these generalization tendencies recover, and we have retroactive inhibition when we measure retention of the first list or proactive inhibition when we measure retention of the second. Forgetting produced in this fashion is added, in a sense, to forgetting stemming from intra-task generalization.

3. *Between a task and other unknown tasks.* When *S* learns a task, it is presumed that part of the process consists in differentiating it from other tasks which he has learned but which are not under control of *E*. With the passage of time, generalization recovers and produces interference at recall.

Such a theory cannot be completely developed until more is known about response generalization. The assumptions concerning the characteristics of response generalization made in Chapter IX will certainly have to have independent empirical verification before the theory can gain a firm footing. Yet at present, such a theory (based on both stimulus and response generalization) seems to account most adequately for the available facts.

SUMMARY

1. *Retention* refers to *amount remembered* of what was learned at some previous time. *Forgetting*, a reciprocal term, refers to *amount not remembered.*

2. Since variables which increase rate of learning retard rate of forgetting, fairly accurate predictions concerning forgetting can be made by studying learning rates.

3. Four response measures are used in studying forgetting: (*a*) recall, (*b*) relearning, (*c*) recognition, (*d*) reconstruction.

4. The relationships between rate of forgetting and each of the following variables were discussed: (*a*) time between learning and retention test; (*b*) degree of learning; (*c*) type of material, including meaningfulness, intra-test similarity, affectivity, and

degree of motor involvement; (*d*) massing vs. distribution of practice; (*e*) context; and (*f*) other learning.

5. Forgetting is believed to be due to interference among response tendencies of different tasks. A theory which accounts for this interference on the basis of generalization was discussed briefly.

CHAPTER XVI

Work

INTRODUCTION

The area of study. The term *work*, as the psychologist uses it, is not easy to define. The most specific delimitation of its meaning which we shall attempt is to confine its use to the performance of *S* when he continues to repeat a well-learned behavior sequence. Actually, the topic work is implicit in all the major behavioral phenomena with which this book has dealt. Repetition of a response may change our discriminal processes (as when visual acuity is lowered); work is vitally dependent upon motivation; work may bring frustration or conflict; work output may vary as a function of transfer of training. Indeed, we shall find that separating work from learning phenomena is a difficult task.

Work, as it shall be thought of here, reduces to an experimental procedure. *S* learns a response as well as possible, so that little facilitation will occur with further practice; the changes which occur thereafter are dealt with as the experimental study of work. The presumption might be that during this stage a decrement in performance will inevitably result. But since we realize that psychological processes are not turned on and off by labelling them with new words, we will not be deluded into thinking that the processes which produce learning cease when the topic changes from learning to work. Therefore, we should not assume that decrements will always occur.

Some of the possible work trends are represented schematically in Fig. 92. First, a learning curve is pictured that has reached an asymptote for a given set of conditions. Thereafter, continued repetition may bring any one of several possible work curves. These curves are depicted as showing decrements, but no one

doubts that with increased motivation, or, with improved work methods (343), the curves could rise. Part of our task in this chapter will be to discuss certain variables which produce differences in rate of work decrement as abstractly shown in the various curves in Fig. 92.

In recent years three books (13, 23, 327) largely concerned with work and work efficiency have been published. These books contain reviews of the literature covering several different ap-

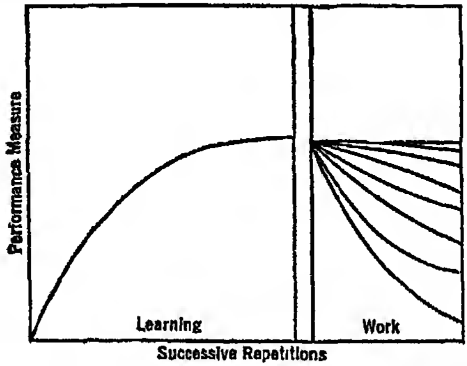


FIG. 92. Schematic relationship of work and learning. After a response is learned, continued repetition of that response may bring a variety of work decrement curves depending upon the manipulation of conditions under which that work is performed.

proaches to the study of work. The three approaches treated most extensively are:

1. *Physiological Studies.* The physiological "cost" of various kinds of work. Experiments determine metabolic rates, circulatory changes, oxygen intake, and so on, as these accompany work.

2. *Industrial Production.* Experiments in work output as a function of different conditions which may be manipulated in the actual factory-type work situation. The investigations are set up to determine how changes in such conditions as lighting, temperature, length of work hours, and so on, affect production.

3. *Psychological Laboratory Studies.* Analytical investigations of work use standardized tasks in an effort to determine specific causal factors influencing work output.

In keeping with the general practice in this book, the

physiological studies will not be taken up. Our major emphasis will be on the third approach but through it we will be able to perceive some of the problems which must be faced in the second approach.

Plan of chapter. In this chapter it will be necessary to stress measurement and interpretation problems. By an interpretation problem is meant the problem of assigning a specific causal factor to account for an obtained result. *E*, of course, in carrying through an experiment, will attempt to control all variables except one. However, in the case of long work periods it is quite possible that other variables do not remain constant, so the problem arises how to interpret the data obtained. In view of the importance of these problems from the analytical point of view, the one we have stressed throughout the book, we shall devote considerable time to them. The major divisions of the chapter are as follows:

1. Tasks used to study work
2. General analysis of work
3. Measurement problems
4. Motivation and rate of work
5. Task variables
6. Transfer of work decrement
7. Rest pauses

TASKS USED TO STUDY WORK

Tasks which have been used in the study of learning have rarely been used in studying work also, apparently because the experimenter cannot control enough of *S*'s time to get him well-practiced before the work experiment is initiated. The most common procedure therefore has been to use tasks which fall into one of two categories. (1) Tasks which, because of their simplicity, show very little practice effect. Simple manual work (such as lifting weights) or color naming are representative. (2) Tasks at which the performance has already reached a high level of proficiency in the normal course of *S*'s life. Such a task is illustrated by simple arithmetic addition. On the whole, laboratory tasks used to study work have been relatively simple as compared with those used to study human learning.

For discussion purposes we may divide work tasks into three categories, depending upon the amount of apparent muscular involvement.

High muscular involvement. Physical labor is studied in the laboratory by using a device called an *ergograph*. The ergograph may be adapted to measure work output of several different muscle groups, such as those of the finger, the arm, or the leg. To operate an ergograph *S* need only pull a weight repeatedly through a given distance. Let us illustrate this by describing a finger ergograph. *S* is seated at a table with his arm strapped to a board. Usually, all fingers, except the one to be used in pulling the weight, are also held firmly against the board. The hand is strapped palm up, and to the free finger a cord is attached. The cord runs over a pulley on the edge of the table. Thus a weight on the end of the cord will rise and fall as *S* flexes and straightens the finger. A small, light stylus attached to the cord marks on slowly rotating kymograph, recording the number of pulls and the distance of each pull. Typically, *S* works at a constant rate, alternately flexing and straightening the finger at a rate determined by a metronome or some other timing device. The instructions usually request *S* to work until he can no longer lift the weight. Occasionally a spring "load" has been used instead of the weight. Some of the variables in manipulating the ergograph are obvious: the amount of load, the distance of the pull, and the frequency of pull per unit of time.

The unit for measuring work output on the ergograph (or on other muscular tasks) may vary. Often, the unit used for the ergograph is a *kilogram-meter*, which is the amount of work required to lift a kilogram weight one meter against gravity. *E* measures the total distance the weight is pulled (from readings on the kymograph) and multiplies this by the size of the weight. He thus secures the total of work done in a fixed time in kilogram-meters. Since the general shape of the work curve is often of interest, the total work period may be divided into several sections of equal duration and the total work done during each section determined.

A somewhat different method of measurement is illustrated by the curve of muscular work in Fig. 93. In this work *E* determined the height of contractions for successive fifteenths of the work

period. The data on which the curve is based are mean values for nine work periods for one *S* who, during each work period, lifted a 24-pound weight with his arm.

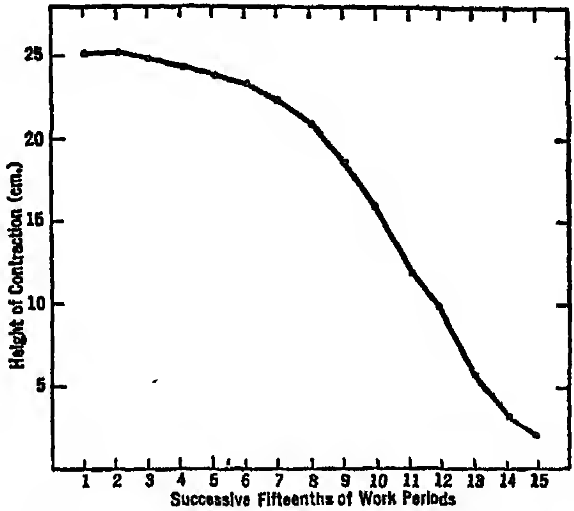


FIG. 93. A work curve for simple muscular contraction in lifting a 24-lb. weight. Data from Neifeld and Poffenberger (280).

Medium muscular involvement. A great many tasks fall roughly in the middle range of a continuum defined as amount of overt muscular involvement. To cite three of these tasks as samples:

1. *Letter Writing. S* is asked to write a certain series of letters (such as *abc*) over and over again as rapidly as possible.

2. *Cancellation Tests. S* is given sheets with many rows of letters appearing on them and is requested to cancel out or draw a line through certain letters, say, *f*, *g*, and *z*, leaving all others uncancelled. The letters may appear as follows:

ghizpxocivubyntmrlqlwkaospftzgriejsutbfhwutodhtjsydgtkziahdzt
gdutospxhfjvkwtquftabzythfitudpqtsdlqpxhbgtrofysftgyokheidtvm, etc.

3. *Assembly-Like Tasks.* Such tasks as sorting articles, boxing or wrapping products, and many others have been used, but usually in the actual factory situation. For example, Rothe has

published a series of articles on work output of women whose major task was to wrap quarter-pound blocks of butter in parchment paper (324).

In most experiments using materials of this nature, *S* works for a constant period of time, the length of the period varying a great deal depending upon what problem *E* is studying and the rate at which he requires *S* to work. The customary response measure is output per unit of time.

Minimum muscular involvement. Again, we list sample tasks falling into this category.

1. *Simple Addition.* This is probably the most frequently used task requiring little overt muscular involvement. The problem may be set up so that *E* pronounces the numbers to be added and *S* responds with the answers, or somewhat more overt motor behavior may be involved if *E* presents the problems on prepared sheets and *S* writes the answers as rapidly as possible.

2. *Color Naming.* This task is the one mentioned as often being used to fill rest intervals in studies of distributed practice. *S* views a board on which a great many colored paper squares are pasted and he is required to name these colors in order as rapidly as possible.

3. *Complex Tasks.* Complex tasks such as *sentence completion tests, mental multiplication,* and even *intelligence tests* have occasionally been used in work studies.

A 10-hour performance curve for simple addition is shown in Fig. 94. Eight *S*s worked from 7:30 A.M. until 5:30 P.M. adding columns of five 2-digit numbers. The problems were presented on work sheets. It is clear that even though *S* was working at a somewhat lower level at the end of the day than he was at the beginning, he was in no sense incapacitated. Note also that errors tended to increase somewhat as the day wore on. The general shape of the curve of correct responses indicates that this is not a simple phenomenon; instead, the output is probably determined by a number of factors which vary from hour to hour.

It is quite obvious that the dimension, degree of motor involvement, we have used in listing sample work tasks does not at all complete the description of the tasks. There are other differences, such as extent of demands for perceptual acuity (cancellation vs. ergographic work), degree of complexity (intelligence test

performance vs. color naming), and homogeneity-heterogeneity (writing *abc* vs. arithmetic addition). Later there will be room for some data on the relationship between rate of work and such dimensions.

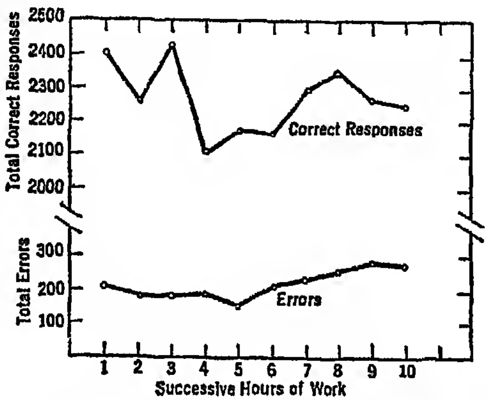


FIG. 94. A 10-hr. work curve in adding numbers. Data from Reed (302).

GENERAL ANALYSIS OF WORK

We have observed work curves and know in general the operations by which such curves are derived. Before introducing the variables which influence the rate of work, hence the shape of the curves, it will be helpful to dip briefly into a general analysis of the work situation. By so doing it will be possible to integrate more adequately the results of the various experiments which follow.

Work inhibition. It has been apparent that in studying work *Es* have chosen tasks for which there is very little learning or practice effect. Or, they have attempted to get *S* well practiced on the task before initiating the series of work trials. Now, it must be asked how one is to view work in relation to learning. Learning, it has been said, is a logical construct; it is a hypothetical process inferred from an observed increment in performance. There is another logical construct applicable when a *decrement* in performance occurs with continued repetition of a response under conditions which might previously have yielded an increment. This construct may be called *work inhibition.*

Characteristics of work inhibition. Three characteristics must be assigned work inhibition if it is to be a useful construct in explaining empirical data.

1. Work inhibition is generated each time a response occurs. This implies that it occurs during learning, i.e., during the very trials in which there is observed an increment in performance. We therefore must assume that when an increment is observed positive effects of learning are greater than negative effects of work inhibition. However, the learning process reaches a limit under a given set of conditions so that further repetition adds but little to it. If it is assumed that work inhibition continues to accumulate with each response, we can see that effectiveness of learning would be reduced, and we may sooner or later expect to encounter a decrement in performance.

2. The amount of work inhibition which is generated is directly related to extensiveness and vigor or effort of musculatures involved in making the response.

3. Finally, it must be assumed that work inhibition dissipates with the passage of time. This is a necessary assumption in view of the common observation that decrements in performance will be reduced if rest periods occur.*

Relationship of work inhibition and previously discussed concepts. 1. In Chapter XI we first discussed the possibility that facilitation by distributed practice (as compared with massed) may be explained by assuming that a negative process is generated each time a response is made, and that this process dissipates more rapidly with the passage of time than does the excitatory process. We can now identify this negative process as work inhibition, as the construct is used here. Empirical data have been presented which show that distribution of practice is more beneficial the greater the overt motor behavior required by the task. Since work inhibition is produced by responding, we may suspect that the greater the motor involvement, the greater is the amount of work inhibition which accrues and the more beneficial rest periods will

* Work inhibition is not used here as a negative motivational state. In the simplified characteristics here assigned work inhibition, we assume only that it reduces the probability of response repetition. See Hull (165) and Solomon (361) for more extensive and somewhat different treatments of work inhibition.

be. We should expect that, in work studies, distribution of practice will greatly facilitate output. This has been confirmed many times.

2. We have also seen that some theorists believe extinction of a conditioned response results from the development of a work inhibition (cf., Mowrer and Jones, 275), and that the greater the amount of work required to make a response the more rapid is the extinction. Note, however, that extinction involves the removal of the incentive (which is not true in work studies) so no one need expect work curves to show characteristics of extinction curves.

3. In the introduction to the chapter on forgetting (XV) it was suggested that processes which cause forgetting are also present during learning, and that these could be thought of as negative processes, which opposed the positive or excitatory processes. The negative processes which cause forgetting are not to be considered work inhibition, since they have characteristics quite the opposite of those demanded of work inhibition. That is, processes which cause forgetting are *reduced* in inhibitory effectiveness during learning and *recover* in effectiveness with the passage of time. We know that if a task is greatly overlearned, by requiring *S* to relearn it many times, very little forgetting takes place. This is the situation which is presumed to exist in the study of work; the task is so well learned that forgetting processes probably play a very minor rôle in the performance. However this may be, we shall assume it in order to reduce the complexity of the analyses which have to be made.

The complexity of work data requires that a conceptual framework be provided for ordering the results of various studies. The above discussion provides the particular point of view to be employed, but this view need not prejudice our evaluation of experimental procedures as such.

MEASUREMENT PROBLEMS

Ideally, work studies would be designed to hold motivation constant and learning at zero. However, psychologists have not deceived themselves by assuming that all learning is removed from work situations or that motivation is always constant. Later

the relationship of motivation and work will come up; at present we need to look into problems pertaining to learning in the work situation.

Many years ago Thorndike (377) fully recognized that the actual decrement in performance produced by what we are calling work inhibition was probably offset to some extent by learning. Thorndike proposed, therefore, that steps be taken to measure the work decrement independent of learning, and sug-

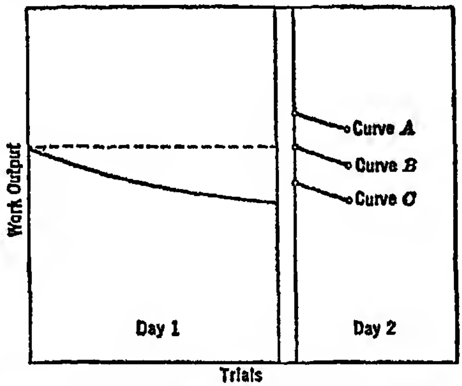


FIG. 95. Illustration of Thorndike's method for measuring work decrement. See text for explanation.

gested a method which we shall call the work-rest method. Figure 95 shows a work curve for Day 1, and three different hypothetical performance curves obtained the following day on the same task. By the work-rest method the amount of actual work decrement produced on Day 1 is determined by the difference in performance at the end of Day 1 and at the start of Day 2. Work inhibition generated on Day 1 dissipates, thus leaving the response strength at the start of Day 2 entirely dependent upon learning. If, for example, the measurement on Day 2 was that shown by Curve A, the work decrement which took place on the previous day must have been much greater than was apparent during that day. The fact that the curve at the start of Day 2 is higher than at the start of Day 1 suggests that some learning took place but was masked by the work inhibition on Day 1. Thus, the total work decrement which actually developed is assumed to be the

difference between the score at the end of Day 1 and the score at the start of Day 2.

If one obtained Curve *B* on Day 2 he would have no evidence of learning, and the measured decrement from beginning to end of work on Day 1 would indicate the effects of work inhibition. If he obtained Curve C he would probably conclude either that not enough time had elapsed for the work inhibition to dissipate or that forgetting had taken place (assuming that motivation was constant on both days).

Figure 94 showed a 10-hour performance curve for simple addition. The performance curve as a whole suggested some decrement, although output was not seriously impaired. In this experiment *S*s returned the following morning for a 10-minute work period in order to allow application of the work-rest method of measuring the work decrement which had taken place on the preceding day. Let us examine the evidence collected. On the last 10 minutes of the first day, *S*s correctly worked 424 problems. On the 10-minute period of the second day, they correctly worked 503 problems. Thus, on Day 1 it appears that the observed work decrement, obscured by practice effects, was less than the "true" decrement. As determined by the work-rest method the influence of work inhibition decreased performance by 79 problems (503–424), which, since there were 8 *S*s, is an average of over 9 problems in a 10-minute period, or about 1 problem per minute.

Difficulty of interpretation. In actual practice, some *E*s have used the work-rest method and some have not. If a stimulus variable is being manipulated, *relative* work decrement produced by the various conditions can be obtained without using the method. If we are interested in determining the absolute influence of work inhibition on output, the work-rest method will probably give a better estimate than will measurements during a single work period. Certain considerations, however, will show that precise interpretation of the obtained measurements is very difficult. To use the work-rest method we assume that motivation is the same at the end of Day 1 as at the start of Day 2. Since it is very difficult to determine the level of motivation at the two points, the equality of motivation must be assumed, and this is indeed a rather tenuous assumption.

"But," you may reply, "isn't decrease in motivation a cause of

work decrement, and, if a difference is produced between Day 1 and Day 2 because of differences in motivation, isn't one justified in using the method?" It will, of course, be seen that decrease in motivation may certainly produce a work decrement. But the rest period between Day 1 and Day 2 is not inserted to allow for an increase in motivation; instead it is there to allow a dissipation of work inhibition which is defined as being independent of motivation. Hence, if the motivation is not constant at the two points it is a question just what we are measuring.

Two other factors complicate the interpretation of results obtained by the work-rest method. If *S* knows that he is about to finish work on Day 1, he may show an *end spurt* in performance, a phenomenon which is believed to be motivational in nature. Secondly, some work curves show a *warm-up effect*, for at the start performance accelerates sharply. This phenomenon, if present, would confuse the interpretation of data taken on Day 2. All things considered, the major asset of using the work-rest method for measuring work decrement is that it demonstrates the complexity of work phenomena and it shows that practice effects may obscure the extensiveness of work decrement. The method, however, can in no sense be thought of as the method which measures "true" work decrement. The problems involved in interpreting data gathered by the work-rest method are not as difficult if very short work periods are used, let us say, 10 to 30 minutes. Fluctuation in motivation is likely to be much less for short intervals than for long intervals. With such short work periods it may be possible to use fairly short rest intervals, although the length of period needed for the dissipation of work inhibition is largely conjectural for tasks involving learning.

It should be pointed out that the work-rest method of measuring work decrement is almost the same method used to measure reminiscence (Chapter XV). The only difference is that a control group is not used in measuring work, the assumption being that the work curve may easily be extrapolated to the next trial if that trial came immediately after the work on the first day. It is not only because the two phenomena are measured in the same fashion that the similarities should be pointed out, but also because both recovery from work and reminiscence (in motor learning, at least) may be one and the same thing. The difference in point of

measurement—reminiscence being measured during learning and work recovery following work decrement i.e., after further improvement by learning is unlikely—should not obscure the possibility that both may depend upon the same mechanism.

This general analysis of measurement problems shows that in dealing with work decrement we shall have difficulty in pointing out specific causal mechanisms which produce the observed performance. Probably no other situation brings together at one time so many antagonistic behavior determinants as does work, this being especially true under prolonged work.

In the design of experiments on work, no new problems are encountered. All four Design Methods have been used, with Design Methods II and IV most common in laboratory studies. In several experiments a single group of *Ss* has worked in all conditions by Design Method IV, and then has repeated the same conditions several times in different orders. Each repetition of all conditions is commonly called a cycle; thus, there may be several cycles in the experiment.

Other measurement problems. Thus far we have said nothing about *Ss'* feelings concerning work. Next to weather notes, casual interchange of conversation probably deals most frequently with topics of personal reactions to work. "I'm tired"; "I'm bored"; "I feel restless," are comments which indicate that we make what are technically discriminations (no matter how flippant) about our work situations. Since nearly everyone works, the personal aspects of work would seem to provide a fertile source of data. At present, however, about all we can safely say concerning the subjective aspects of work is that there *are* such aspects, and that they vary among individuals and within the same individual at different times and in different work situations. Bartley and Chute (13), in their recent book, use the term *fatigue* to indicate the subjective, personal feelings which are experienced by *S* in a work situation. Such feelings, according to these authors, are the most pertinent data in the work situation, and they believe that when techniques or methods are developed by which these experiences can be measured for each individual, the current emphasis on work output will become secondary in the laboratory. For present purposes, however, we shall have to retain the emphasis upon work output.

MOTIVATION AND RATE OF WORK

Introduction

In long term experiments on work it is almost impossible to segregate the influence of motivation from the influence of other factors. This problem is present in learning investigations but to a less critical extent. In the first place, most learning experiments cover relatively short intervals; extended learning experiments contain provisions for rest intervals, so that motivation will not drop. Furthermore, in studying learning, motivation can be separated most clearly from other factors by using animals. Animals, however, can seldom be made to do extended work because performance must be accompanied constantly by incentives which in turn reduce the motivation. It is possible that the larger animals, such as horses, might be capable of becoming laboratory subjects, but the white rat would be difficult to handle in a work situation. We do not say, of course, that motivation of humans is not reduced by long periods of work, but everyone knows that humans will continue to work for long periods. It is known in such cases, therefore, that motivation is not zero.

Preliminary analysis. Let us suppose that *S*, working under a given set of conditions, shows a gradual decrement in his work output. We offer him a sizeable prize for increasing his performance. He does so, and the curve rises. Are we to say that this demonstrates learning? From a sheer definitional standpoint it is learning, although it is not commonly called such unless the performance rises to a level beyond that which it had been at the start of the work series. Yet, this criterion is not satisfactory. It is possible that the rise in performance beyond previous levels reflects learning which had taken place during the interval between the time when the work period started and the introduction of the incentive.

The rate of rise is probably the most satisfactory criterion we have for distinguishing between new learning as a consequence of additional incentives and old learning (which is depressed by work inhibition and only shows itself when an incentive is introduced). We know that learning is a gradual process. If, following the introduction of an incentive in a work situation, the performance "jumps" to a higher level, it is unlikely that new learning

is involved. Such a situation was reviewed in Chapter VI, and the results shown in Fig. 32. There was no evidence of gradual improvement; rather, it was as though the incentive (competition) brought out ability already acquired. On the other hand, there is evidence that new learning can be involved in such situations. In Kitson's classic study (199), hand compositors in a printing firm were offered a bonus for more rapid work. The men had been employed in this capacity for an average of 10 years before introduction of the bonus system. For many weeks after introduction of the bonus, production continued to rise; there was no jump to a given level. If this were merely a matter of the incentive "bringing out" previously acquired learning, we would have expected a sharp rise in output. Such evidence leads to the conclusion that incentives offered for increased rate of work may not only increase the motivation to exhibit learning which had already taken place, but may also provide a facilitating condition for more learning.

The fact that incentives may increase rate of work demonstrates the possibility that increased motivation counteracts work inhibition. Let us examine in some detail several experiments relating to incentive variation and set.

Incentive Variation

Positive incentives have been shown universally to (1) delay work decrement, (2) prevent it, or (3) increase performance. These relationships hold only if the incentive is pertinent to a motivating condition. If a young girl working in a factory is forced to turn her pay check over to her father each week, a money bonus would probably bring little improvement in her production, whereas, let us say, a bonus of a fur coat would. The incentive must be such as to enhance and sustain a dominant motive.

Ergographic work. In an experiment by Preston, Brotemarkle, and Campbell (299), 48 male *S*s were used to determine the influence of simple incentives on work output of both the finger and the arm. All *S*s served in 2 sessions, 1 week apart. An unusual technique was used to arrange experimental and control conditions, in that incentives were given for work with the left arm and left finger but no incentives were given for work with the

right arm and right finger. On the first session, all *S*s worked with no special incentives in producing an ergogram for their right arms and fingers, and left arms and left fingers. On the second session, ergograms were taken on the right arm and right finger without incentives and on the left arm and left finger with incentives. Differences in the amount of work at the first and second session with right arm and finger as compared with differences in work with left arm and finger gave the quantitative measure of the influence of the incentives. The order of work was counter-balanced for the different appendages during the first session. On the second session the ergograms on the right arm and right finger were taken before those on the left arm and finger. Thus, there was no possibility of transfer of incentive influence, and if there were any transfer of work decrement from the right arm and right finger work to the left arm and left finger work, this procedure would tend to minimize the influence of the incentives.

After taking the ergograms on the right arm and right finger on the second session, *E* told *S*: (Remember, the *S*s were men).

> Last week I studied the records you made in a number of ways and discovered that you were doing just a little better on the whole than the average of 100 women studied here last year. I think that you can do better than that. What I am going to do is to give you either some cigarettes or some gum provided your record indicates that you are doing both of two things, i.e., pulling harder on each pull and making more pulls (299, p. 499).

After these instructions the ergogram on the left arm was taken. Note that the instructions actually might increase more than one motive. In them *S* is compared unfavorably with women; he is told that he should be able to do better, and finally offered an incentive, although the cigarettes and gum were probably quite insignificant as compared with the unfavorable comparison with women. After taking the ergogram, *E* gave *S* the gum or cigarettes and told him that he had improved, and asked if he would try to do even better on the next ergogram, which was for the left finger.

The results of this experiment are shown in Table 32. The values indicate the mean differences between the first session and control session for each appendage. The number of pulls is used as one measure and the total height of the pulls in millimeters is used as the other. All differences between the first and second

session for the right arm and right finger are insignificant; all differences between the left arm and left finger for the two sessions are highly significant statistically. Thus, we may conclude, the instructions which E gave S produced a significant difference in the work output.

TABLE 32

INFLUENCE OF INCENTIVES ON ERGOGRAPHIC WORK

		Mean Difference Between First and Second Sessions	
		Number Pulls	Distance Pulled (mms.)
No Incentives	Right Arm	1.25	150.0
	Right Finger	2.37	10.4
Incentives	Left Arm	4.31	1092.0
	Left Finger	10.30	126.0

Data from Preston, Brotemarkle, and Campbell (299).

Dynamometer work. The *dynamometer*, you will remember, is a device for measuring strength of grip. Occasionally it has been used to study work, with each grip consituting a trial. Knowledge of performance, you know, may enhance performance. Knowledge of performance may facilitate because it gives information concerning the correct response, and because it serves as a motivating factor. In the experiment now to be described there was no correct response, so we may consider the influence of knowledge of performance as being entirely motivational in nature. That is, S's own performance may serve as an incentive for improvement.

This experiment, by Manzer (244), was carried out with 68 men and 68 women, each group being divided into two sub-groups to serve as experimental and control *Ss*. On each trial the control groups were asked to grip the dynamometer as hard as they could. This was repeated 50 times at 15-second intervals. At no time did the control *Ss* know how well or how poorly they did. Experimental *Ss*, on the other hand, worked for the first 10 trials in the same manner as the control *Ss*, but for trials 11 through 30, E

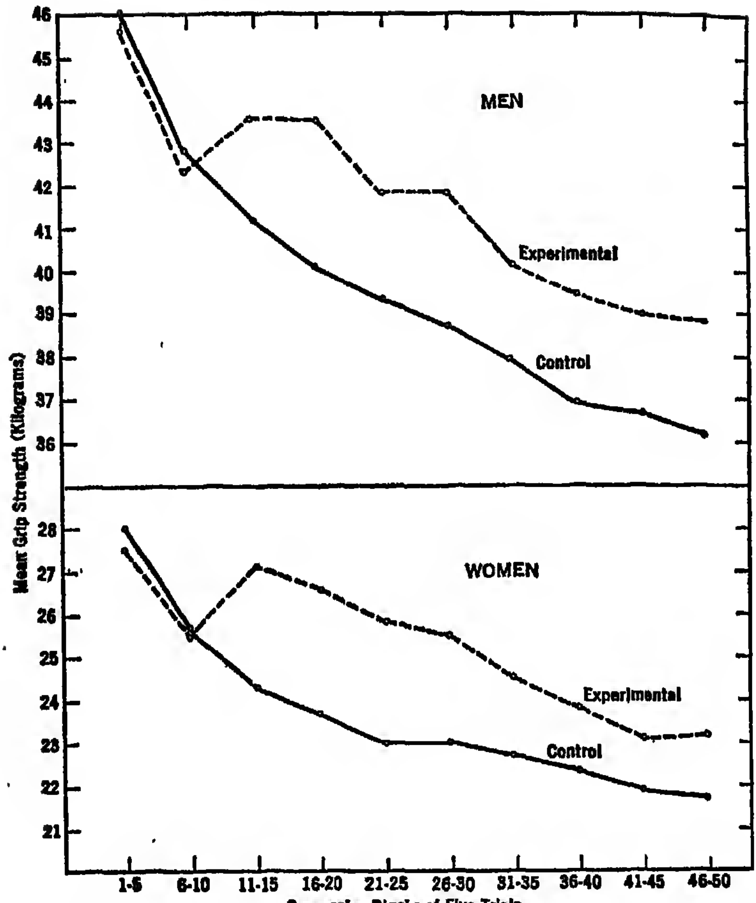


Fig. 96. The influence of knowledge of performance on work. Knowledge of performance was given after trials 11 through 30 for the experimental groups. Data from Manzer (244).

told them how much they had gripped (in kilograms) after each attempt. From trial 31 on through trial 50, the information was withheld. The important data are the differences between the two groups, starting with the eleventh trial. These results are plotted in Fig. 96, by combining successive blocks of 5 trials. The groups were well matched on the initial 10 trials so that differences

occurring after that point must be attributed to knowledge of performance. Influence of the knowledge of performance is marked. Performance is high initially and remains high even after the information is removed.

Perceptual-motor task and social facilitation. Unless the task is a highly complex one, the presence of an audience usually facilitates work. Indeed, if *S* has been working for some time on an ergograph and seems near collapse, work output will increase if onlookers are brought into the room. There is also some evidence that working in a group will facilitate performance. An experiment by Mukerji (279), a British investigator, will demonstrate this.

Mukerji used two tasks, one a cancellation task in which *S* was asked to cross out only T's and X's from among scores of letters on a page. The other task may be called the letter-shape task. In this task *S* was presented sheets on which capital letters were shown running in zigzag lines all over the sheets. *S* was to place a C under all curved letters (O, P, B, etc.) and S under all straight letters (A, K, L, etc.). The 31 *S*s had been given considerable preliminary training on these tasks, so that the practice effect on the work series would be small. There were 18 boys and 13 girls, 11 to 14 years of age. Initially all *S*s worked in a group on each task for 5 minutes, with instruction to do as much as possible in the time allowed. Then, each *S* worked in isolation on the tasks for the same amount of time. If there were a practice effect, it would tend to facilitate work in isolation. A rest period between group work and isolated work allowed for dissipation of any inhibition which may have developed during the group session.

TABLE 33

Effect of Working in Groups vs. Working in Isolation

Work Condition	*Task*	*Mean Score*
Group	Cancellation	149
	Letter-Type	160
Isolated	Cancellation	114
	Letter-Type	132

Data from Mukerji (279).

The results (Table 33) show that group work output was superior to isolated work. The group presumably provides a form of competition, which we have seen to be one way of increasing motivation (Chapter VI).

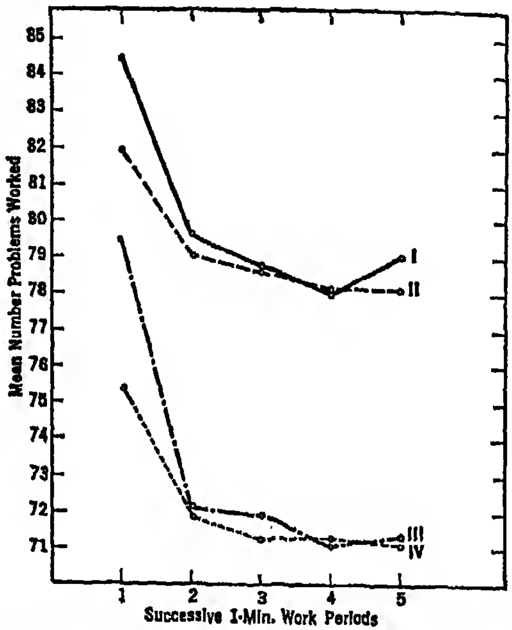


FIG. 97. Initial work rate as a function of time and amount sets. Data from Krueger (207).

Set

We have consistently treated set as an aspect of motivation. Sets determine which stimuli *S* attends to, the goals he strives for, and in general, constitute the directive function of motivation as discussed in Chapter VI. One type of experiment which has been performed concerns work output as related to the amount of work which *E* specifies must be done (thereby inducing an amount set).

As an illustration, consider four conditions of an experiment by Krueger (207). The task involved adding series of 2 one-place numbers; all 36 *S*s served in all conditions, by a form of systematic

randomization, following 6 practice days. Before each condition Krueger instructed his *S*s concerning the amount of work or the length of the work period which they were to have on that particular day. The 4 conditions were as follows:

Condition I: *S* told that he would work for 5 minutes.
Condition II: *S* told that he would work 500 problems.
Condition III: *S* told that he would work for 50 minutes.
Condition IV: *S* told that he would work 4000 problems.

Thus, *S* was faced with different amounts of work at the start of each period. The influence of these different sets is most clearly seen in the initial trials as demonstrated in Fig. 97. The unit of measurement was the number of problems actually completed during a minute, and the graph shows the mean output for the first 5 minutes. One general principle is clear: the greater the amount of work which *S* has to do, the less is the initial output. *S* no doubt brings some past experiences to such a situation. The instructions, in effect, tell him that with a great deal to do he had better not "burn himself out" by working too hard initially.

Set and Distraction

Distraction experiments are those in which *E* attempts to break *S*'s attention or set by forcing him to attend to irrelevant stimuli. There are no general conclusions which can be drawn concerning these experiments. Some have shown that distracters will reduce performance, and others have shown that they will not. Much apparently depends on the task, intensity of distracting stimuli, length of time experienced, and whether or not *S* has preconceived ideas as to whether distraction should or should not influence performance. It has been demonstrated, however, that even when distraction does not inhibit performance, the energy expended is greater under the distraction condition than under the non-distraction condition (269). Such findings suggest that under distraction there may be compensatory motivation which functionally isolates *S* from distraction, although how this is accomplished is unknown.

One of the most unusual studies of this problem is Hovey's (156). He chose to give intelligence tests to *S*s. With such a task there is no continual repetition of a single sequence, but we do

have a work situation, since the practice effect is slight. Equivalent forms of the intelligence test were used, the first form for matching the groups and the second for one group under distraction and for another under quiet. There was no systematic variation in Hovey's distracters; instead, he simply had all manner of unusual stimuli present. There were buzzers buzzing at irregular intervals; bells ringing; organ pipes blowing; a phonograph playing popular music at full blast; a spotlight swinging around; men marching around in the room above the test room; men walking and talking in the test room, and so on. In short, it was a chaotic situation for 20 minutes. Nevertheless, under these conditions, there was only a slight decrement produced in the scores on the intelligence test.

A study by Laird (208), however, has shown that for a simulated assembly job, noise does have an influence on production. *S* stood in front of a moving tape in which many holes were punched, and inserted a stylus through each hole to contact a metal plate which activated a counter. *S* attempted to keep up with the tape by inserting the stylus through all holes, which meant 10,500 insertions in a half day. With such a task, and with long periods of work, Laird found definite differences in production as a function of intensity, complexity, and pitch of the sounds which were used as the experimental variables. As intensity, predominant pitch, and complexity of noises increased, the errors (number of omissions) increased. A complex sound, varying in intensity and pitch, had the most serious effect on work.

Although there are no available data on the matter, it seems likely that the interest in a task (which is a form of motivation) will be important in determining whether or not presumed distractions do distract. The observation that extraneous stimuli may distract us while we are studying but not while we are reading a detective story is probably valid.

Baker's experiment. An important factor which must be considered in experiments on distraction and work has been discovered by Baker (7). He hypothesized that whether or not distraction stimuli retard performance is a function in part at least of *S*'s preconceived notions concerning the influence of the distracting stimuli. To test the hypothesis Baker attempted to

give *S*s certain sets concerning the influence of extraneous stimuli. The task was the addition of numbers. *E* gave *S* a two-place number and asked him to add successively 6, 7, 8, and 9 to this number, and report the successive answers. Thus, if *E* gave the number 32, *S* started by adding 6 to this number (and reporting 38 to *E*), then adding 7 (and reporting 45), then 8, then 9. After 9, *S* would immediately start over again by adding 6 to the number, then on up to 9 and repeat. For each two-place number given, *S* added as accurately and as rapidly as possible for 30 seconds. At the end of this interval another two-place number was given and again *S* would add 6, 7, 8, and 9 as rapidly as possible. The number of figures added during a 30-second interval became the basic response measure.

Baker used 7 groups of *S*s, but his method and results for only three of these groups will be reviewed. Each group worked 10 days working under 2 conditions each day, one a distraction condition, the other a non-distraction (control) condition. Each condition on each day consisted of ten 30-second work periods. On the first day *S* was given the distraction condition first and the non-distraction second; on the second day the order of the 2 conditions was reversed, and at the end of 10 days each condition had come first 5 times.

The control conditions were the same for all three groups—all *S*s merely added for 5 minutes. The experimental conditions for the three groups were as follows:

Group I: *S*s were told that the experiment concerned the influence of distraction on work. Each *S* was shown his previous day's work curves under both distraction and non-distraction conditions.

Group II: *S*s were told that the experiment concerned the influence of distraction on work. Furthermore, they were shown a graph of work curves which were misrepresented to them as genuine. These faked curves indicated that in a previous experiment other *S*s had performed *better* under distraction than under non-distraction conditions. As in Group I, *S*s in Group II were also shown their work curves from day to day.

Group III: *S*s were treated exactly the same as those in Group II except that the fictitious curves demonstrated that *S*s had performed *more poorly* under distraction than under normal conditions.

The distraction which Baker provided was a phonograph record playing dance music at an intensity level of about 70 decibels (roughly equal to the intensity of noise from several typewriters being operated in the room). *S*s were given no practice days, so that actually there was some improvement in performance from

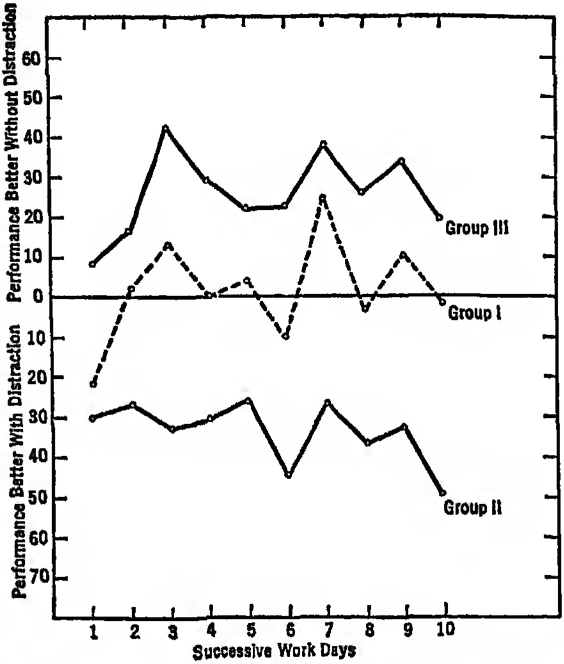


FIG. 98. The influence of set on work under distraction. Data from Baker (7).

day to day. The major data are the differences between performance with and without distraction. These differences for the three groups are plotted in Fig. 98. The zero line indicates no difference between distraction and non-distraction. In general, *S*s of Group I showed no better performance under one condition than under the other. *S*s of Group II, on the other hand, did consistently better with distraction, and Group III did consistently better without distraction.

Baker's results show that instructions which give *S* an expectation of what his performance may be actually produce a shift of the

performance in that direction. The cause for this is undoubtedly motivational in nature. The raw data show that the differences plotted in Fig. 97 are made up of shifts not only in the experimental conditions but also in the control conditions. The performance of Group II on the control condition was considerably lower than that of Group III on the same condition. Assuming that the groups were equal in ability, these results imply that the motivation was different for the 2 conditions.

It is Baker's opinion that in cases where *E*s have found differences in work under distraction and non-distraction (which he did not find for Group I), it is likely that *S*s have entered the experiment with pre-conceived ideas of the influence of distraction. Whether such sets will account for the influence of distraction for long work periods (such as the several hours used by Laird) is questionable. Laird's finding of systematic changes in output as a function of systematic changes in noise does not preclude such an explanation as suggested by Baker, but Baker's technique should be tried out for work periods of extended length.

Conclusion

Motivation is probably the most important single factor governing rate of work. By varying incentives, marked changes in work output can be produced. Sets toward amount of work to be done and the influence of sets on distraction further show the importance of motivation. The real possibility of shifts in motivation with extended work spells make interpretation of decrements in terms of other variables very shaky. It will be necessary to consider further interpretative problems of this kind in the following sections.

TASK VARIABLES AND RATE OF WORK

In studying learning we were able to distinguish different task dimensions with which rate of learning was associated. Because of the continuity of learning and work, we might suspect that the variables important for learning would also be important for work. But the fact is that there are very few systematic data on the matter. Furthermore, it is very difficult to obtain such data.

If E manipulates a task variable, such as meaningfulness, it is difficult for him to tell whether or not motivation has also varied for extended work spells. As for showing that rate of work is different for grossly different kinds of tasks, that is quite simply done. The problem arises in interpreting the differences, i.e., are the differences due to the nature of the task as such, or are they due to the fact that interest (motivation) varies as a function of the task?

Let us illustrate this interpretative problem with an experiment. Poffenberger (294) studied work decrement with different tasks, among which was a task in simple addition and an intelligence test. *S*s worked on both tasks for 5 hours with results showing an actual increment in performance on the intelligence test and a decrement in addition. What produced the difference? Was it that more work inhibition accrued in addition, or was it that the variety of problems presented on the intelligence test sustained motivation? Or both?

In addition to this problem, we still have the difficulty of obtaining equality for all task dimensions except one, e.g., when we vary intra-task similarity, meaningfulness must be held constant. Recognizing that interpretation of work as a function of different task variables may be ambiguous, let us inquire briefly into the specific relationships, and then return to the interpretation problem.

Intra-Task Similarity

Depending upon its nature, intra-task similarity appears to have two somewhat different relationships to rate of work. In work studies only formal similarity has been varied; you will remember that formal similarity refers to degree of identity of parts of a task. Formal similarity has been varied in tasks which require little discrimination for adequate performance and in tasks which require considerable discrimination. Consider an illustration of each of these situations.

Repetition with little discrimination. Robinson and Bills (314) used the letter-writing task and varied the number of elements to be written. In one condition S wrote *ab ab ab* over and over again as rapidly as possible for 20 minutes; in another condition he wrote *ab cd ef* over and over again as rapidly as possible for

the same period of time. At the end of each minute during the work period *E* signaled *S* to make a distinguishing mark on his paper. Thus, *E* was able to determine the number of two-letter units written per minute under each of the two conditions.

From our general theoretical analysis it would follow that the condition in which *ab* was written would show greater decrement than the one in which *ab cd ef* was repeated. Work inhibition should accumulate more rapidly in the first instance because of the more exact repetition of response. The results of the experiment substantiate the expectation.

Robinson and Bills discovered an interesting correlate of work decrement in letter-writing, namely, an increase in the size of the letters written. As work progresses the size of the letters increases until it approaches the size normally characteristic of children. Those of you who have written long essay examinations under pressure of time will probably confirm this finding. It was Robinson's hypothesis (313) that as the fine musculature used in writing develops work inhibition there is a gradual shift to the more gross musculatures of the arm. Because these muscles are less well practiced in carrying on this particular skill, the precision of writing suffers.

The Robinson-Bills experiment has shown that the greater the degree of identity of successive work units the more rapid is the decrement. No one can infer from this experiment that writing *ab* does not influence the writing of, let us say, *cd*; that is, there is undoubtedly some overlapping of the response mechanisms in writing the two units, and the writing of one will to a certain extent influence the other. What the experiment does demonstrate is that when the overlapping is 100 per cent, a greater decrement will result than if the overlapping is partial. Because the work period in this experiment was fairly short, and because the basic nature of two tasks is quite similar, one would not expect differences in motivation to account for the obtained results.

Repetition with discrimination. Color-naming is a task in which variation in formal identity produces results quite the opposite from those produced in letter-writing. Bills (22) required *Ss* to name colors as rapidly as they could for 10 minutes. One condition involved 2 colors on the color board, another 3, and another 5.

Although the decrement was not great for any condition the amount was directly related to the number of colors, i.e., naming 2 colors produced less decrement than naming 5. On the surface, at least, these results seem contradictory to the letter-writing experiment reported above. In color-naming the less the repetition of a given response the greater is the work decrement, whereas in letter-writing the greater the repetition of a given response the greater is the decrement. Perhaps the difference is due to the discrimination which is required in the color-naming situation, and these discriminations cannot be made as rapidly among 5 colors as between 2. However, results from another experiment by Bills and Robinson (314) cloud the over-all relationship. In this study letters were placed in random order on a board, 100 to a board, and *S* was requested to name them as colors are named from a color board. Five different conditions were distinguished by the number of different letters on the boards, these being 2, 4, 8, 16, and 24 letters. With this situation the amount of work decrement decreased as the number of letters increased, thus showing the same relationship as in writing letters and just the opposite from that shown in naming colors. What would produce such different results in naming colours vs. naming letters from a board is not clear, and to date no one has offered an adequate explanation for the discrepancies. It is possible that our greater familiarity with numbers, as compared to colors, would reduce intra-task interference so that work decrement from repetition becomes the major factor. On the basis of the data, however, we must conclude that the relationship between repetition with discrimination and work decrement is ambiguous.

Other Possible Task Variables

We have already assumed that the greater the degree of muscular involvement the more rapid is the work decrement. This assumption was made even though only partial evidence exists for such a principle. It is clear, for example, that work decrement on an arm ergograph is much more rapid than that for simple addition. Furthermore, at the level of gross observation it seems clear that such tasks differ markedly in muscular involvement. What we do not know, however, is whether degree of muscular

involvement is the pertinent difference between the tasks causing differences in work decrement. This we only assume.

There is some evidence that meaningfulness may be a variable (315), but there are contingent variables which have been inadequately analyzed so that we are unable to say that meaningfulness is the pertinent one. It is probable that in long work spells interest in the task will be directly related to degree of meaningfulness, so that it will probably be necessary to use very short work periods in further experimentation on this variable, or follow the scheme suggested below.

A Suggested Approach for Investigation of Task Variables

The most serious objection we have discussed with reference to long work spells as a function of type of material concerns differential motivation. Now, let us ask if there is a way by which this obstacle can be removed so that the interpretation of studies can be less ambiguous.

The starting point may be in an empirical evaluation of the intrinsic motivating characteristics of the tasks. By this we mean that for the average *S* certain tasks elicit more interest than do others, and this might be quantified in order to obtain tasks which vary in certain aspects but not in terms of their interest rating. Certain experiments on work satiation give us a lead which might be worth following. In a work satiation experiment *S*s are given a task and told they are to keep working until they are "fed up" with it. *S*s are not given a speed set but are merely told to keep at the task. Let us take an illustration.

Burton instructed his *S*s as follows:

> This is an experiment to study certain aspects of motor behavior. Your task is to draw simple figures called "moon-faces". You can see that it consists of a face or circumference, and two eyes and a nose represented by three dots. Do not correct a figure once you have drawn it.
>
> On the signal begin drawing a series of "moon-faces" on this paper. There are pencils here for your use. Now this is important: Continue drawing *moon-faces until you are fed up and have had enough of it.* When that time comes you may stop.
>
> From time to time I will give you a short rest period. I will let you know when that time comes. Speed is not a factor in this experiment and this is not a test of persistence. Do you understand? (51, p. 324).

Burton's results show that individuals vary considerably in the length of time they will continue to work at such a task. On the average, these college sophomores worked 17 minutes, 36 seconds, and on the average drew 392 moon-faces.

Suppose we gave these *S*s other tasks to do with the same instructions? Would we expect that the length of time they worked to vary as a function of the task? Direct evidence on this is lacking. However, in one of Poffenberger's experiments (293) involving addition of numbers, *S*s were told that they were to work until they refused to go on. This was a work experiment, and *S*s were given no rest periods as they were in Burton's satiation experiment. Nevertheless, the shortest period worked by any one of the 10 *S*s was 71 minutes. One *S* worked for 285 minutes (nearly 5 hours). Burton's results, taken in conjunction with those of Poffenberger suggest that we *might* be able to measure the average incentive value of a given task by submitting *S*s to satiation experiments under non-work conditions. That is, we would give instructions similar to those used by Burton and use the length of time worked as a measure of the motivating power of the task. Having obtained such motivating values for several different tasks, we might then proceed to determine the influence of other variables by which tasks differ. For example, say that we have two tasks which have equal motivating value but differ widely in intra-task similarity. Differences in work curves for extended work spells would not then be attributed to differences in motivation, but to differences in intra-task similarity.

Satiation experiments provide an operational definition of boredom. Boredom, as currently interpreted (cf. Barmack, 11, and Bartley and Chute, 13), reflects a simple conflict situation. Motivation to work at the task is provided by the instructions, by *S*'s desire not to be a "quitter," by his desire to please *E*, and so on. As he works a conflicting motive to quit the task develops. We presume that the original motives for work are reduced, too. At the point where the motive to stop becomes stronger than those to continue, he is satiated. Boredom increases as the difference in the strength of the opposed motives decreases.

Other concomitants of satiation experiments might be useful for equating motivating power of tasks. When the task is very simple, *S* does many things to attempt to vary the response. Moon-

faces become many different characters, and little changes in the position of the eyes and nose will be noted. *S*s who are forced to name colors for extended periods may vary the pitch of their voices over a wide range, or they may actually attempt to sing the names of the colors.

The importance of these variations is that they seem to amuse *S*, or keep him interested longer in an otherwise nearly intolerable situation. With more complex tasks, such as addition, many different combinations of numbers are given, so that variation in behavior is automatically required by the work. This analysis suggests that complexity of task, defined in terms of diversity of response required to perform it, is an extremely important variable if for no other reason than it sustains motivation. When we increase the complexity of a task we not only decrease work inhibition (because of forced variation of response to the complex task), but we also increase the likelihood of sustaining motivation. Both factors would tend to reduce observed work decrement.

The analytical approach which we have suggested would discover whether or not these two variables are inevitable co-variables. Must motivation always vary as task variables (meaningfulness, intra-task similarity, complexity, and so on) are manipulated? By using satiation experiments we may be able to equate empirically the motivation-evoking properties of tasks. If this can be done with tasks which vary along other dimensions, we can proceed to determine whether or not work decrement varies as a function of these other dimensions. The control of motivation still remains the most difficult problem in analytical experiments. It should be repeated, that we may observe work performance over long periods under given conditions, and obtain data having general significance for practical work situations, but the scientist who is in search of specific causal factors finds such studies quite unsatisfying. Analytical studies are necessary before he can state relationships between type of task and work decrement.

One final comment concerning motivation. It has been continually stressed that it is difficult to interpret work studies because we lack control of motivation over extended work spells, or, in lieu of this, we lack ability to measure its changes.

The possibility of doing the latter should not be overlooked. Since our basis for inferring changes in motivation is measured changes in tension, we might well ask why we shouldn't measure the tension changes. Problems concerning such measurements were discussed in Chapter VI. The major drawback at present is our mechanical inability to measure tension adequately for extended periods, especially if overt activity is required in performing the task. Ryan (327, Chapter 5) has a thorough discussion of these problems to supplement the brief treatment given in our Chapter VI.

TRANSFER OF WORK DECREMENT

If transfer of training in the acquisition of tasks is dependent largely on the similarity relationships between tasks, we shall expect the same in the case of transfer of work inhibition. As in learning studies, we have work situations in which the similarity relationships may be carefully specified and others in which only general similarity between tasks can be detected. First, let us consider an illustration of the analytical type of study.

Specific Similarity Relationships

A study by Bills and McTeer (24) has convincingly demonstrated the importance of formal identity between responses in producing work decrement. They used the letter-writing task with the following 5 conditions:

		Standard Task	*Alternate Task*
Cond. I	(all elements common)	*abc*	*abc*
Cond. II	(2 elements common)	*abc*	*abd*
Cond. III	(1 element common)	*abc*	*afe*
Cond. IV	(no elements common)	*abc*	*def*
Cond. V	(control)	*abc*	(*rest*)

Note that Condition I required *S* to repeat *abc* in both the standard and alternate task, and in Conditions II through IV there was less and less identity in the responses of the two tasks. On Condition V, *S* merely rested during the time that was occupied with work on the alternate task in the other conditions. (A consideration of this design will show that it is simply an elaboration of the experiment by Robinson and Bills, previously reviewed,

in which intra-task similarity was varied with *S* writing *ab* in one condition and *ab cd ef* in another.) The 20 *S*s alternated between the two tasks, spending 1 minute on standard tasks, then 1 minute on alternate tasks, for a total of 16 periods. All *S*s went through three cycles of the experiment, i.e., repeated all conditions three times, but the order of the conditions was different on each cycle.

TABLE 34

TRANSFER OF WORK DECREMENT AS A FUNCTION OF SIMILARITY OF TASKS

Condition	*Mean Number Units Written in 8 Min. on Standard Task*
I	395.55
II	412.60
III	420.15
IV	430.45
V	437.80

Data from Bills and McTeer (24).

The results of the experiment are shown in Table 34 in terms of the mean number of three-letter units of the standard task written per cycle. These data show that, from Condition I to V there is less and less work decrement. Indeed, Condition V shows the benefits which are derived from distributed practice (compared with results of Condition I). We may tentatively conclude that the greater the degree of similarity between two tasks the greater will be the work decrement.

General Transfer

Studies of general transfer effects have been concerned with the development of so-called fatigue tests. The question is: "What influence does work on one task have on performance on another?" A partial answer will be found in an illustrative experiment which determined the influence of prolonged automobile driving on performance on a series of standardized tests (326). A common transfer of training design was used with experimental and control conditions set up as follows:

	9:00 A.M.		7:00 P.M.
Experimental:	Tests	.... Drive 300 Miles	Tests
Control:	Tests	....	Tests

Only 6 *Ss* were used, but each served 120 days in the experimental condition and 28 days in the control condition. The 300-mile drive was a standard course, including both rural and urban traffic conditions. The statistical comparison involved the difference in scores on pre- and post-tests for experimental and control conditions. Several tests were used, including color-naming, addition, visual efficiency, and steadiness.

Results were positive. On almost all tasks significant differences were obtained. Steadiness was decreased, visual efficiency was lowered, color-naming and addition were retarded. *If we assume* that *Ss* did not have preconceived sets as to what the influence of the driving should be, we may conclude that the obtained results were a function of the driving. It might be presumed that the lowered test scores reflected lowered driving efficiency, although no evidence was gathered on this matter.

Tests of general fatigue will be maximally useful from a scientific point of view when the relationships between the work task and the tests are known. Until then all that can be said is that it is possible to measure changes in test performance as a consequence of work on another task. That work inhibition will transfer is apparent; the mechanisms by which such transfer is effected cannot at present be inferred from general transfer studies.

REST PAUSES

Rest pauses in work almost always increase the output per unit of time worked, and since the methodology is clear we need not report a study merely to demonstrate the principle. With tasks involving heavy muscular work, rest pauses are obviously necessary if work is to continue. For example, if *S* is requested to work an ergograph at a rate of one pull per second, and if the weight is heavy, he will be exhausted after a short period of time. In order to do any more work he must have some rest. Exhaustion, in a sense, is a forced rest, necessary for survival. Rest pauses will facilitate most of those tasks which require extensive motor behavior. In general, it will be remembered, this is the same principle which is at work in learning as a function of massed and distributed practice. The greater the motor involvement in a task the greater is the benefit of rest pauses.

The considerations above apply only to the sheer work output. It is another matter to consider the subjective feelings of *S* as a function of rest pauses vs. no rest pauses. The fact that *S* can work a full day with little decrement in adding numbers does not mean that subjectively he feels the same at the end of the day. Nor does it mean that he is physically unchanged. The major deterrent to work decrement is increased motivation, and since increased motivation involves increased tension one may expect physiological expenditure to be greater.

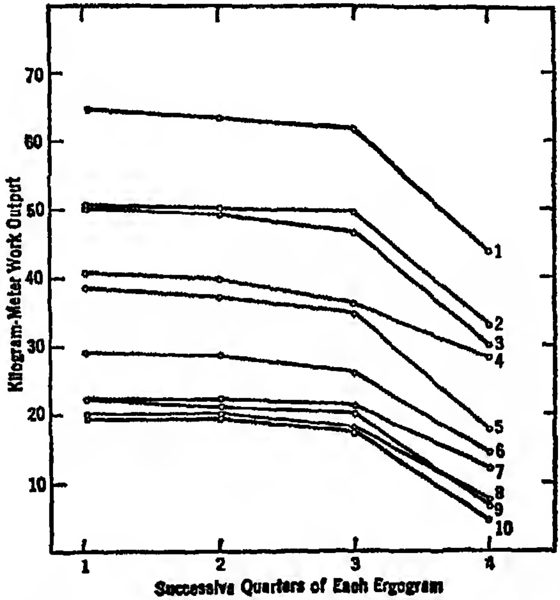


FIG. 99. Decrement shown in each quarter of ten ergograms taken 10 min. apart. Data from Manzer (243).

Recovery from Ergographic Work

To illustrate the relationships between work and length of rest pauses, we choose a study by Manzer (243) which has shown that with the usual ergographic work, recovery from work inhibition is nearly complete after 20 minutes. This varies, of course, as a function of the size of the weight and rate of pull. Of greater interest, however, is the cumulative decrement in

work when the work inhibition is not allowed to dissipate completely before more work is done. Let us examine the results for one *S*.

This *S* worked 10 times on an arm ergograph, each time working until, apparently, he was exhausted. Each trial was followed by a 10-minute rest. The load lifted was 5.5 kilograms and the results are figured in kilogram-meter work units. Each ergogram was divided into 4 equal time spaces, *E* determining the amount of work done in each quarter. For this *S*, the development of decrement is particularly clear from ergogram to ergogram, and within each ergogram (Fig. 99). Presumably, of course, with continued work *S* would have reached a point where he could no longer make a single lift after the 10-minute rest. As it is, there is a clear picture of accumulative decrement, and it can be inferred that the 10-minute rest was not enough to allow all work inhibition to dissipate.

Blocks

In factory or office work it is commonly noted that if rest pauses are not officially provided by management, workers will usually take rest periods at their own discretion. With the usual type of work, there is no reason to believe that such rest periods are necessary to the continuation of work, although work may be more efficient as a consequence of them. In a laboratory study of high speed work for short intervals, Bills (21) has discovered a phenomenon which functionally appears to provide the same benefit as do rest intervals in more extended work of slower speed.

Bills demonstrated this phenomenon with several types of material including color-naming, substitution tasks, and the naming of opposites, but we shall take up method and results for rapid arithmetic addition. *S*s, presented with long columns of figures, alternatively added three to a figure and then subtracted three from the next number and so on down the columns. As 10 *S*s worked continuously 7 minutes a day for 8 days, *E* recorded the time for each response, and whether each response was correct or incorrect. In examining the records Bill discovered that from time to time there was a period of no response, the length of this period occupying time normally required for 2 to 6

responses. Further investigation showed that frequency of these *blocks* was inversely related to work output. This does not mean, of course, that the blocks caused the work decrement, because in effect the blocks were the evidence of work decrement. Data for the 7-minute work for the 8 days combined are shown in Table 35. It is clear from these data that there is some decrement in output during the 7-minute work period and an increase in the number of blocks per minute.

TABLE 35

WORK DECREMENT AND BLOCKS IN HIGH-SPEED ADDITION

Response Measures (*means*)	*Successive Minutes of Work*						
	1	2	3	4	5	6	7
Number Responses	61.0	57.0	56.7	55.7	55.2	55.0	55.8
Number Blocks	2.9	3.4	3.7	3.8	4.2	3.9	4.2
Length Block (sec.)	1.9	2.2	2.3	2.4	2.5	2.9	2.7

Data from Bills (21).

Data considered individually for the 8 successive days show that performance increases from day to day and frequency of blocks decreases slightly. Bills believes that blocks serve functionally as safeguards against impairment, and the reason that marked decrements are not found in tasks with low motor involvement is because the blocks serve as forced brief rest intervals, thus preventing the decrement. It should be kept in mind that this is an hypothesis.

CONCLUDING CONSIDERATIONS

We have made no attempt to come to grips with numerous practical work problems which appear in the factory situation. It may be pertinent in concluding this chapter to mention briefly the nature of some of these problems.

The starting point for investigations of work is the output curve for an "average" day. Typically, such curves show that production does vary rather markedly at different hours of the day, although an attempt to draw any curve as representative would be misleading. The absolute output represented by the

curves is presumed to be a function of many conditions, such as policies of a factory with respect to rest periods, the nature of the tasks, the work payment plan (piece work, hourly, bonus), the rôle of local unions, lighting and ventilation, length of work day, adequacy of supervision, morale in general, and so on. We can well see how difficult it would be to control all variables so that one could say that conditions of optimal efficiency had been attained (cf., Ryan, 327, for a discussion of difficulties in determining optimal efficiency).

There are many other problems which are directly or indirectly concerned with work. Time and motion studies presume to discover the minimum number and kind of movements to produce maximum output with least physical effort. Such studies, however, get at only one small part of a vast complex of conditions which affect work. Turnover rate of employees is sometimes used as a gross index of work conditions within a plant. The effort is made to keep turnover rate at a minimum by manipulation of conditions such as those listed above. Then there is the problem of accident prevention. Attempts have been made to discover conditions to which frequency of accidents is related. Since precision of movements is basic to accident prevention in many jobs, causes of variability in performance become vital in accident investigation and prevention.

Recent trends in industry have emphasized the rôle of the adjustment of the individual as a major source of work data. There is some reason to believe that the output of a worker, his accident rate, the amount of spoilage, and so on, may reflect his personal social adjustment, e.g., whether or not his marriage is a happy one. Indeed, it is possible that future developments will find psychiatric aid as common in large concerns as is medical aid today. The influence and contribution of such aid to working efficiency is as open to experimental investigation as is the influence of amount of light at a work bench.

SUMMARY

1. In this chapter we have been concerned with performance decrements which may occur when *S* continues to repeat a well-learned behavior sequence.

2. Many different tasks have been used to study work in the laboratory. The present discussion of these tasks was organized around a dimension defined as amount of apparent muscular involvement. Ergographic work illustrates high muscular involvement, whereas simple addition or color-naming illustrates low muscular involvement.

3. Work inhibition was introduced as a logical construct to parallel the logical construct of learning. Learning is inferred when an increment in performance is observed, and work inhibition is inferred when a decrement in performance is observed. It was assumed that learning and work inhibition may both develop whenever a response sequence is repeated. Several assumptions about the characteristics of work inhibition were made.

4. The measurement problems discussed were concerned with segregating the influence of learning from the influence of work inhibition, and the segregation of the influence of work inhibition from the influence of motivation. Throughout the chapter we have stressed the difficulty of making clear interpretations of causal factors in work decrement because of the uncontrolled and unknown fluctuations in motivation over long work spells. A technique for equating the intrinsic motivating power of tasks was proposed. This technique would determine the satiation ratios of various tasks as an indication of the motivating power of the tasks.

5. Discussion of variables influencing the rate of work was on a rather general level, the variables considered being motivation, type of task, transfer, and rest pauses.

6. Work brings together in a complex manner all the major psychological processes which we have discussed in the text. Much more research is needed before we will be able to assess the effect of each of these processes.

References

1. ADAMS, C. R. Individual differences in behavior resulting from experimentally induced frustration. *J. Psychol.*, 1940, *10*, 157-176.
2. ALLPORT, G. W. The ego in contemporary psychology. *Psychol. Rev.*, 1943, *50*, 451-478.
3. ALPER, Thelma G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, *41*, 403-420.
4. AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, *37*, 393-411.
5. ANSBACHER, H. L. Distortion in the perception of real movement. *J. exp. Psychol.*, 1944, *34*, 1-23.
6. ARNOLD, Magda B. Physiological differentiation of emotional states. *Psychol. Rev.*, 1945, *52*, 35-48.

7. BAKER, K. H. Pre-experimental set in distraction experiments. *J. gen. Psychol.*, 1937, *16*, 471-488.
8. BARKER, R. G. An experimental study of the resolution of conflict by children. In *Studies in personality*. McNemar, Q., and Merrill, Maud A., Eds. New York: McGraw-Hill, 1942, 13-34.
9. BARKER, R. G. An experimental study of the relationship between certainty of choice and the relative valence of the alternatives. *J. Pers.*, 1946, *15*, 41-52.
10. BARKER, R. G., DEMBO, Tamara, and LEWIN, K. Frustration and regression. An experiment with young children. *Univ. Ia. Stud. Child Welf.*, 1941, *18*, No. 386.
11. BARMACK, J. E. Boredom and other factors in the physiology of mental effort: an exploratory study. *Arch. Psychol., N. Y.*, 1937, *31*, No. 218.
12. BARTLETT, F. C. *Remembering: a study in experimental and social psychology.* London: Cambridge University Press, 1932.
13. BARTLEY, S. H., and CHUTE, ELOISE. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.
14. BECK, S. J. Introduction to the Rorschach method, a manual of personality study. *Monogr. Amer. Orthopsychiat. Ass.*, 1937, No. 1.
15. BELLAK, L. The concept of projection. *Psychiatry*, 1944, *7*, 353-370.

16. Berger, C. I. Stroke-width, form and horizontal spacing of numerals as determinants of the threshold of recognition. *J. appl. Psychol.*, 1944, *28*, 208-231.
17. Berger, C. II. Stroke-width, form and horizontal spacing of numerals as determinants of the threshold of recognition. *J. appl. Psychol.*, 1944, *28*, 336-346.
18. Bergmann, G., and Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, *51*, 1-24.
19. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, *10*, 173-197.
20. Biel, W. C., and Force, R. C. Retention of nonsense syllables in intentional and incidental learning. *J. exp. Psychol.*, 1943, *32*, 52-63.
21. Bills, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, *43*, 230-245.
22. Bills, A. G. Some causal factors in mental blocking. *J. exp. Psychol.*, 1935, *18*, 172-185.
23. Bills, A. G. *The psychology of efficiency.* New York: Harper, 1943.
24. Bills, A. G., and McTeer, W. Transfer of fatigue and identical elements. *J. exp. Psychol.*, 1932, *15*, 23-36.
25. Birch, H. G. The effect of socially disapproved labeling upon a well-structured attitude. *J. abnorm. soc. Psychol.*, 1945, *40*, 301-310.
26. Bird, C. *Social psychology.* New York: D. Appleton-Century, 1940.
27. Bitterman, M. E., and Soloway, E. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, *36*, 134-136.
28. Blodgett, H. C. The effect of introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, *4*, 113-134.
29. Blodgett, H. C., and McCutchan, K. Place versus response learning in the simple T-maze. *J. exp. Psychol.*, 1947, *37*, 412-422.
30. Bobbitt, J. M. An experimental study of the phenomenon of closure as a threshold function. *J. exp. Psychol.*, 1942, *30*, 273-294.
31. Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: D. Appleton-Century, 1942.
32. Bousfield, W. A., and Mote, F. A., Jr. The construction of a tilting activity cage. *J. exp. Psychol.*, 1943, *32*, 450-451.
33. Bray, C. W. Transfer of learning. *J. exp. Psychol.*, 1928, *11*, 443-467.
34. Briggs, L. J., and Reed, H. B. The curve of retention for substance material. *J. exp. Psychol.*, 1943, *32*, 513-517.
35. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, *25*, 323-332.
36. Brogden, W. J. Non-alimentary components in the food-reinforcement of conditioned forelimb-flexion in food-satiated dogs. *J. exp. Psychol.*, 1942, *30*, 326-335.
37. Brogden, W. J. Tests of sensory pre-conditioning with human subjects. *J. exp. Psychol.*, 1942, *31*, 505-517.
38. Brogden, W. J. Sensory pre-conditioning in human subjects. *J. exp. Psychol.*, 1947, *37*, 527-539.
39. Brogden, W. J., and Culler, E. Experimental extinction of higher-order responses. *Amer. J. Psychol.*, 1935, *47*, 663-669.
40. Brown, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, *31*, 272-292.
41. Brown, J. S. The generalization of approach responses as a function of

stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, *33*, 209-226.

42. Brown, W. The judgment of difference. *Univ. Calif. Publ. Psychol.*, 1910, *1*, No. 1.

43. Brown, W. To what extent is memory measured by a single recall? *J. exp. Psychol.*, 1923, *6*, 377-382.

44. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, *16*, 343-361.

45. Bruner, J. S., and Goodman, Cecile C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, *42*, 33-44.

46. Bugelski, R. Extinction with and without sub-goal reinforcement. *J. comp. Psychol.*, 1938, *26*, 121-133.

47. Bugelski, R., and Miller, N. E. A spatial gradient in the strength of avoidance responses. *J. exp. Psychol.*, 1938, *23*, 494-505.

48. Bunch, M. E. Transfer of training in the mastery of an antagonistic habit after varying intervals of time. *J. comp. Psychol.*, 1939, *28*, 189-200.

49. Bunch, M. E. A comparison of retention and transfer of training from similar material after relatively long intervals of time. *J. comp. Psychol.*, 1941, *32*, 217-231.

50. Bunch, M. E., and McCraven, V. G. The temporal course of transfer in the learning of memory material. *J. comp. Psychol.*, 1938, *25*, 481-496.

51. Burton, A. Behavioral characteristics of monotony in two age groups. *J. exp. Psychol.*, 1943, *33*, 323-329.

52. Burtt, H. E., and Falkenberg, D. R., Jr. The influence of majority and expert opinion on religious attitudes. *J. soc. Psychol.*, 1941, *14*, 269-278.

53. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, *2*, No. 2.

54. Buxton, C. E. Retroaction and gains in motor learning: III. Evaluation of results. *J. gen. Psychol.*, 1940, *22*, 309-320.

55. Buxton, C. E. The frame of reference. In *Fields of psychology*. Seashore, R. H., Ed. New York: Holt, 1942.

56. Buxton, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, *40*, 313-340.

57. Buxton, C. E. Interaction of meaningfulness and length of rest interval as factors in measured reminiscence. *Amer. Psychol.*, 1948, *3*, 235-236.

58. Buxton, C. E., and Newman, E. B. The forgetting of "crowded" and "isolated" materials. *J. exp. Psychol.*, 1940, *26*, 180-198.

59. Carter, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, *28*, 481-490.

60. Cason, H. Backward conditioned eyelid reactions. *J. exp. Psychol.*, 1935, *18*, 599-611.

61. Cofer, C. N. An analysis of errors made in the learning of prose materials. *J. exp. Psychol.*, 1943, *32*, 399-410.

62. Cofer, C. N. and Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, *49*, 513-540.

63. Cofer, C. N., Janis, Marjorie, and Rowell, Mary M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, *32*, 266-269.
64. Cook, S. W. The production of "experimental neurosis" in the white rat. *Psychosom. Med.*, 1939, *1*, 293-308.
65. Cook, T. W. Studies in cross education: I. Mirror tracing the star-shaped maze. *J. exp. Psychol.*, 1933, *16*, 144-160.
66. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, *41*, 330-355.
67. Cook, T. W. Amount of material and difficulty of problem solving: II. The disc transfer problem. *J. exp. Psychol.*, 1937, *20*, 288-296.
68. Courts, F. A. The knee-jerk as a measure of muscular tension. *J. exp. Psychol.*, 1939, *24*, 520-529.
69. Courts, F. A. Relations between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, *25*, 235-256.
70. Courts, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, *39*, 347-367.
71. Cowles, J. T. Food-tokens as incentives for learning by chimpanzees. *Comp. Psychol. Monogr.*, 1937, *14*, No. 71.
72. Crafts, L. W. Transfer as related to number of common elements. *J. gen. Psychol.*, 1935, *13*, 147-158.
73. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, *55*, 467-517.
74. Culler, E., Finch, G., Girden, E., and Brogden, W. Measurements of acuity by the conditioned-response technique. *J. gen. Psychol.*, 1935, *12*, 223-227.

75. Dashiell, J. F. Affective value-distances as a determinant of esthetic judgment-times. *Amer. J. Psychol.*, 1937, *50*, 57-67.
76. Dashiell, J. F. *Fundamentals of general psychology*. Cambridge, Mass.: Houghton Mifflin, 1937.
77. Davis, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, *23*, 141-158.
78. Davis, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, *39*, 329-346.
79. Denny, M. R. The rôle of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, *36*, 373-389.
80. Dietze, A. G., and Jones, G. E. Factual memory of secondary school pupils for a short article which they read a single time. *J. educ. Psychol.*, 1931, *22*, 586-598.
81. Dollard, J., *et al. Frustration and aggression.* (Second printing) New Haven: Yale University Press, 1939.
82. Dulsky, S. G. The effect of a change of background on recall and relearning. *J. exp. Psychol.*, 1935, *18*, 725-740.

83. Ebbinghaus, H. *Memory: A contribution to experimental psychology.* Trans. by Ruger, H. A., and Bussenius, Clara E. New York: Bureau of Publications, Teachers College, Columbia University, 1913.
84. Edwards, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm soc. Psychol.*, 1941, *36*, 34-50.

85. EDWARDS, A. L. The retention of affective experiences—a criticism and restatement of the problem. *Psychol. Rev.*, 1942, *49*, 43-53.

86. EDWARDS, A. L. A critique of "neutral" items in attitude scales constructed by the method of equal appearing intervals. *Psychol. Rev.*, 1946, *53*, 159-169.

87. EISENBERG, P., and CHINN, H. A. Tonal range and volume level preferences of broadcast listeners. *J. exp. Psychol.*, 1945, *35*, 374-392.

88. ELLIOTT, M. H. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, *4*, 19-30.

89. ELLSON, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, *23*, 339-358.

90. ERICKSEN, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, *31*, 339-345.

91. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, *57*, No. 263.

92. EYSENCK, Margaret D. An experimental and statistical study of olfactory preferences. *J. exp. Psychol.*, 1944, *34*, 246-252.

93. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, *38*, 111-131.

94. FESTINGER, L. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *J. exp. Psychol.*, 1943, *32*, 291-306.

95. FESTINGER, L. Studies in decision: II. An empirical test of a quantitative theory of decision. *J. exp. Psychol.*, 1943, *32*, 411-423.

96. FINGER, F. W. Quantitative studies of "conflict": I. Variations in latency and strength of the rat's response in a discrimination-jumping situation. *J. comp. Psychol.*, 1941, *31*, 97-127.

97. FINGER, F. W. Convulsive behavior in the rat. *Psychol. Bull.*, 1947, *44*, 201-248.

98. FOLEY, J. P., Jr., and COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, *32*, 168-175.

99. FREEMAN, G. L. A method of inducing frustration in human subjects and its influence upon palmar skin resistance. *Amer. J. Psychol.*, 1940, *53*, 117-120.

100. FRENCH, Elizabeth G. Avoidance behavior in conflict situations. Ph.D. dissertation, Yale University, 1942.

101. FRENCH, J. R. P. Authority and frustration. Studies in topological and vector psychology III: Organized and unorganized groups under fear and frustration. *Univ. Ia. Stud. Child Welf.*, 1944, *20*, 231-308.

102. GAGNÉ, R. M., FOSTER, Harriet, and CROWLEY, Miriam E. The measurement of transfer of training. *Psychol. Bull.*, 1948, *45*, 97-130.

103. GARRETT, H. E. Variability in learning under massed and spaced practice. *J. exp. Psychol.*, 1940, *26*, 547-567.

104. GAUGER, Marguerite E. The modifiability of response to taste stimuli in the pre-school child. *Teach. Coll. Contr. Educ.* 1929, No. 348.

105. GHISELLI, E. E., and BROWN, C. W. Learning in accident reduction. *J. appl. Psychol.*, 1947, *31*, 580-582.

106. GIBSON, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, *47*, 196-229.

107. GIBSON, Eleanor J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, *28*, 93-115.

108. GIBSON, Eleanor J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, *30*, 185-200.

109. GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1041, *38*, 781-817.

110. GIBSON, J. J., JACK, Eleanor G., and RAFFEL, Gertrude. Bilateral transfer of the conditioned response in the human subject. *J. exp. Psychol.*, 1932, *15*, 416-421.

111. GILBERT, R. W. A further study of the effect of non-informative shock upon learning. *J. exp. Psychol.*, 1937, *20*, 396-407.

112. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, *35*, 255-269.

113. GODBEER, Elizabeth. Factors introducing conflict in the choice behavior of children. M.A. thesis, Yale University, 1940.

114. GOLDEN, C. S., and RUTTENBERG, H. J., *The dynamics of industriol democracy.* New York: Harper, 1942.

115. GRAHAM, C. H., and GAGNÉ, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, *26*, 251-280.

116. GRAHAM, D. T. Experimental transfer of conditioning in dogs. *J. exp. Psychol.*, 1944, *34*, 486-493.

117. GRAHAM, Frances K. Conditioned inhibition and conditioned excitation in transfer of discrimination. *J. exp. Psychol.*, 1943, *33*, 351-368.

118. GRANT, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, *35*, 393-402.

119. GRANT, D. A., and NORRIS, Eugenia B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, *36*, 390-397.

120. GRANT, D. A., NORRIS, Eugenia B., and BOISSARD, Suzanne. Dark adaptation and the pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1947, *37*, 434-439.

121. GRAYBIEL, A., CLARK, B., and MACCORQUODALE, K. The illusory perception of movement caused by angular acceleration and by centrifugal force during flight: I. Methodology and preliminary results. *J. exp. Psychol.*, 1947, *37*, 170-177.

122. GREENBERG, Pearl J. Competition in children: an experimental study. *Amer. J. Psychol.*, 1932, *44*, 221-248.

123. GRICE, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, *41*, 137-143.

124. GRICE, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, *38*, 1-16.

125. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

126. GULLIKSEN, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, *53*, 199-213.

127. GURNEE, H. The effect of electric shock for right responses on maze learning in human subjects. *J. exp. Psychol.*, 1938, *22*, 354-364.

128. GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

129. GUTHRIE, E. R. *Psychology of human conflict.* New York: Harper, 1938.

130. GUTHRIE, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, *43*, 1-20.

131. GUTHRIE, E. R., and HORTON, G. P. *Cats in a puzzle box.* New York: Rinehart, 1946.

132. HAAGEN, C. H. Scaling the principal attributes of words to be used in experiments in verbal learning. M.A. thesis, State University of Iowa, 1942.

133. HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Ph.D. dissertation, State University of Iowa, 1943.

134. HALL, W. E., and CUSHING, J. R. The relative value of three methods of presenting learning material. *J. Psychol.*, 1947, *24*, 57-62.

135. HAMILTON, G. V. A study of trial and error reactions in mammals. *J. Anim. Behav.*, 1911, *1*, 33-66.

136. HAMILTON, G. V. A study of perseverance reactions in primates and rodents. *Behav. Monogr.*, 1916, *3*, No. 13.

137. HAMILTON, J. A., and KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, *16*, 237-253.

138. HAMILTON, R. Jane. Retroactive facilitation as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1943, *32*, 363-376.

139. HANAWALT, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol., N. Y.*, 1937, *31*, No. 216.

140. HARLOW, H. F. The formation of learning sets. Presidential Address, Midwestern Psychol. Ass., 1948.

141. HARLOW, H. F. Studying animal behavior. In *Methods of psychology.* Andrews, T. G., Ed. New York: Wiley, 1948.

142. HARRIS, J. D. Forward conditioning, backward conditioning, pseudo-conditioning, and adaptation to the conditioned stimulus. *J. exp. Psychol.*, 1941, *28*, 491-502.

143. HASLERUD, G. M. Frustration as an experimental problem: III. Some interrelations of behavioral measures of frustration in chimpanzees. *Character and Pers.*, 1938, *7*, 136-139.

144. HEIDBREDER, Edna. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, *35*, 191-223.

145. HEIDBREDER, Edna. Studying human thinking. In *Methods of psychology.* Andrews, T. G., Ed. New York: Wiley, 1948.

146. HEIDBREDER, Edna, and OVERSTREET, Phoebe. The attainment of concepts: V. Critical features and contexts. *J. Psychol.*, 1948, *26*, 45-69.

147. HENRY, N. B., Ed. The psychology of learning. Part II of *Forty-First Yearbook, Nat. Soc. Study Educ.* Bloomington, Ill.: Public School Publishing, 1942.

148. HERON, W. T., and SKINNER, B. F. Changes in hunger during starvation. *Psychol. Rec.*, 1937, *1*, 51-60.

149. Hevner, Kate. An empirical study of three psychophysical methods. *J. gen. Psychol.*, 1930, *4*, 191-212.

150. Hilgard, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, 1938, *35*, 282-297.

151. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.

152. Hilgard, E. R., and Marquis, D. G. Acquisition, extinction, and retention of conditioned lid responses to light in dogs. *J. comp. Psychol.*, 1935, *19*, 29-58.

153. Hilgard, E. R., and Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.

154. Hilgard, E. R., and Smith, M. B. Distributed practice in motor learning: score changes within and between daily sessions. *J. exp. Psychol.*, 1942, *30*, 136-146.

155. Honzik, C. H. The sensory basis of maze learning in rats. *Comp. Psychol. Monogr.*, 1936, *13*, No. 64.

156. Hovey, H. B. Effects of general distraction on the higher thought processes. *Amer. J. Psychol.*, 1928, *40*, 585-591.

157. Hovland, C. I. Experimental studies in rote-learning theory: I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, *22*, 201-224.

158. Hovland, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, *23*, 172-190.

159. Hovland, C. I. Experimental studies in rote-learning theory: V. Comparison of distribution of practice in serial and paired-associate learning. *J. exp. Psychol.*, 1939, *25*, 622-633.

160. Hovland, C. I. Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, *26*, 568-587.

161. Hovland, C. I. Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, *27*, 271-284.

162. Hovland, C. I., and Sears, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, *23*, 477-493.

163. Hull, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, *39*, 25-43.

164. Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, *45*, 730-734.

165. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

166. Hull, C. L. Reactively heterogeneous compound trial-and-error learning with distributed trials and terminal reinforcement. *J. exp. Psychol.*, 1947, *37*, 118-135.

167. Hull, C. L., *et al. Mathematico-deductive theory of rote learning.* New Haven: Yale University Press, 1940.

168. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, *25*, 141-158.

169. HUMPHREYS, L. G. Distributed practice in the development of the conditioned eyelid reaction. *J. gen. Psychol.*, 1940, *22*, 379-385.

170. HUMPHREYS, L. G. Extinction of conditioned psycho-galvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, *27*, 71-78.

171. HUMPHREYS, L. G. The strength of a Thorndikian response as a function of the number of practice trials. *J. comp. Psychol.*, 1943, *35*, 101-110.

172. HUNT, J. McV. The effects of infant feeding-frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.*, 1941, *36*, 338-360.

173. HUNT, W. A., and VOLKMANN, J. The anchoring of an effective scale. *Amer. J. Psychol.*, 1937, *49*, 88-92.

174. HURLOCK, Elizabeth B. An evaluation of certain incentives used in school work. *J. educ. Psychol.*, 1925, *16*, 145-159.

175. HURLOCK, Elizabeth B. The use of group rivalry as an incentive. *J. abnorm. soc. Psychol.*, 1927, *22*, 278-290.

176. HUSBAND, R. W. A note on maze learning with the time factor constant. *J. gen. Psychol.*, 1929, *2*, 366-369.

177. HUSBAND, R. W. Analysis of methods in human maze learning *J. genet. Psychol.*, 1931, *39*, 258-278.

178. HUSBAND, R. W. Comparative behavior on different types of mazes, *J. gen. Psychol.*, 1931, *5*, 234-244.

179. IRWIN, F. W. The realism of expectations. *Psychol. Rev.*, 1944, *51*, 120-126.

180. IRWIN, O. C. Effect of strong light on the body activity of newborns. *J. comp. Psychol.*, 1941, *32*, 233-236.

181. JENKINS, W. O., and SHEFFIELD, F. D. Rehearsal and guessing habits as sources of the "spread of effect." *J. exp. Psychol.*, 1946, *36*, 316-330.

182. JOHNSON, D. M. A modern account of problem solving. *Psychol. Bull.*, 1944, *41*, 201-229.

183. JOHNSON, W. Language and speech hygiene. An application of general semantics. *Gen. Sem. Monogr.*, 1939, No. 1.

184. JONES, F. N., and JONES, Margaret H. Vividness as a factor in learning lists of nonsense syllables *Amer. J. Psychol.*, 1942, *55*, 96-101.

185. JONES, H. E. Trial and error learning with differential cues. *J. exp. Psychol.*, 1945, *35*, 31-45.

186. JOST, H. Some physiological changes during frustration. *Child Develpm.*, 1941, *12*, 9-15.

187. KARN, H. W. Sensory pre-conditioning and incidental learning in human subjects. *J. exp. Psychol.*, 1947, *37*, 540-544.

188. KEISTER, Mary E., and UPDEGRAFF, Ruth. A study of children's reactions to failure and an experimental attempt to modify them. *Child. Develpm.*, 1937, *8*, 241-248.

189. Keller, F. S., and Schoenfeld, W. N. Studies in International Morse Code: III. The efficiency of the code as related to errors made during learning. *J. appl. Psychol.*, 1944, *28*, 254-266.

190. Keller, F. S., and Taubman, R. E. Studies in International Morse Code: II. Errors made in code reception. *J. appl. Psychol.*, 1943, *27*, 504-509.

191. Kellogg, W. N. The learning curve for flying an airplane. *J. appl. Psychol.*, 1946, *30*, 435-441.

192. Kellogg, W. N., and Walker, E. L. "Ambiguous conditioning," a phenomenon of bilateral transfer. *J. comp. Psychol.*, 1938, *26*, 63-77.

193. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, *36*, 212-220.

194. Kendler, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, *40*, 265-270.

195. Kendler, H. H., and Underwood, B. J. The rôle of reward in conditioning theory. *Psychol. Rev.*, 1948, *55*, 209-215.

196. Kientzle, Mary J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, *36*, 187-211.

197. Kimble, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, *37*, 1-15.

198. Kingsley, H. L. *The nature and conditions of learning.* New York: Prentice-Hall, 1946.

199. Kitson, H. D. A study of the output of workers under a particular wage incentive. *Univ. J. Business*, 1922, *1*, 54-68.

200. Klebanoff, S. G. An experimental analysis of approach-approach and avoidance-avoidance conflict. M.A. thesis, Yale University, 1939.

201. Klee, J. B. The relation of frustration and motivation to the production of abnormal fixations in the rat. *Psychol. Monogr.*, 1944, *56*, No. 257.

202. Kleemeier, R. W. Fixation and regression in the rat. *Psychol. Monogr.*, 1942, *54*, No. 246.

203. Koch, S., and Daniel, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, *35*, 167-187.

204. Kohler, W. *The mentality of apes.* New York: Harcourt, Brace, 1925.

205. Krueger, W. C. F. The effect of overlearning on retention. *J. exp. Psychol.*, 1929, *12*, 71-78.

206. Krueger, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, *17*, 145-153.

207. Krueger, W. C. F. The influence of amount limits and time limits upon the rate of works. *J. appl. Psychol.*, 1937, *21*, 113-118.

208. Laird, D. A. The influence of noise on production and fatigue, as related to pitch, sensation level, and steadiness of the noise. *J. appl. Psychol.*, 1933, *17*, 320-330.

209. Lanier, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, *11*, 199-217

210. Lantz, Beatrice. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, *59*, No. 271.

211. LEAVITT, H. J., and SCHLOSBERG, H. The retention of verbal and of motor skills. *J. exp. Psychol.*, 1944, *34*, 404-417.

212. LEEPER, R. W. A motivational theory of emotion to replace "emotion as disorganized response." *Psychol. Rev.*, 1948, *55*, 5-21.

213. LEUBA, C., and LUCAS, C. The effects of attitudes on descriptions of pictures. *J. exp. Psychol.*, 1945, *35*, 517-524.

214. LEVINE, R., CHEIN, I., and MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, *13*, 283-293.

215. LEVINE, J. M., and MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, *38*, 507-517.

216. LEWIN, K. Environmental forces. In *A handbook of child psychology*. Murchison, C., Ed. Worcester, Mass: Clark University Press, 1933.

217. LEWIN, K., DEMBO, Tamara, FESTINGER, L., and SEARS, Pauline S. Level of aspiration. In *Personality and the behavior disorders*. Vol. I. Hunt, J. McV., Ed. New York: Ronald Press, 1944.

218. LEWIN, K., LIPPITT, R., and WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, *10*, 271-299.

219. LIDDELL, H. S. Conditioned reflex method and experimental neurosis. In *Personality and the behavior disorders*. Vol. I. Hunt, J. McV., Ed. New York: Ronald Press, 1944.

220. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.*, 1932, *22*, No. 140.

221. LINDAHL, L. G. Movement analysis as an industrial training method. *J. appl. Psychol.*, 1945, *29*, 420-436.

222. LONG, L., and WELCH, L. Influence of levels of abstractness on reasoning ability. *J. Psychol.*, 1942, *13*, 41-59.

223. LOUCKS, R. B. The experimental delimination of neural structures essential for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1935, *1*, 5-44.

224. LOCKS, R. B., and GANTT, W. H. The conditioning of striped muscle responses based upon faradic stimulation of dorsal roots and dorsal columns of the spinal cord. *J. comp. Psychol.*, 1938, *25*, 415-426.

225. LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, *54*, No. 248.

226. MCCLELLAND, D. C. Factors influencing the time error in judgments of visual extent. *J. exp. Psychol.*, 1943, *33*, 81-95.

227. MCCLELLAND, D. C., and APICELLA, F. S. A functional classification of verbal reactions to experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, *40*, 376-390.

228. MCCLELLAND, D. C., and ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, *25*, 205-222.

229. MCCORD, F. The effect of frustration on hoarding in rats. *J. comp. Psychol.*, 1941, *32*, 531-541.

230. MCGARVEY, Hulda R. Anchoring effects in the absolute judgement of verbal materials. *Arch. Psychol., N. Y.*, 1943, *39*, No. 281.

231. McGeoch, J. A. The influence of associative value upon the difficulty of nonsense-syllable lists. *J. genet. Psychol.*, 1930, *37*, 421-426.

232. McGeoch, J. A. Forgetting and the law of disuse. *Psychol. Rev.*, 1932, *39*, 352-370.

233. McGeoch, J. A. *The psychology of human learning*. New York: Longmans, Green, 1942.

234. McGeoch, J. A., and Melton, A. W. The comparative retention values of maze habits and of nonsense syllables. *J. exp. Psychol.*, 1929, *12*, 392-414.

235. Maier, N. R. F. Reasoning in white rats. *Comp. Psychol. Monogr.*, 1929, *6*, No. 29.

236. Maier, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, *10*, 115-143.

237. Maier, N. R. F. Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, *12*, 181-194.

238. Maier, N. R. F. An aspect of human reasoning *Brit.. J. Psychol.*, 1933, *24*, 144-155.

239. Maier, N. R. F. *Studies of abnormal behavior in the rat*. New York: Harper, 1939.

240. Maier, N. R. F. The behavior mechanisms concerned with problem solving. *Psychol. Rev.*, 1940, *47*, 43-58.

241. Maier, N. R. F. Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, *35*, 349-360.

242. Maier, N. R. F., and Longhurst, Joan U. Studies of abnormal behavior in the rat: XXI. Conflict and "audiogenic" seizures. *J. comp. physiol. Psychol.*, 1947, *40*, 397-412.

243. Manzer, C. W. An experimental investigation of rest pauses. *Arch. Psychol., N. Y.*, 1927, *14*, No. 90.

244. Manzer, C. W. The effect of knowledge of output on muscular work. *J. exp. Psychol.*, 1935, *18*, 80-90.

245. Marchetti, P. V. Time-errors in judgments of visual extents. *J. exp. Psychol.*, 1942, *30*, 257-261.

246. Marinho, Heloísa. Social influence in the formation of enduring preferences. *J. abnorm. soc. Psychol.*, 1942, *37*, 448-468.

247. Marks, E. S. Skin color judgments of negro college students. *J. abnorm. soc. Psychol.*, 1943, *38*, 370-376.

248. Marquis, D. P. A study of frustration in newborn infants. *J. exp. Psychol.*, 1943, *32*, 123-138.

249. Martin, R. F. "Native" traits and regression in rats. *J. comp. Psychol.*, 1940, *30*, 1-16.

250. Masserman, J. H. Psychobiologic dynamism in behaviour. *Psychiatry*, 1942, *5*, 341-347.

251. Meehl, P. E., and MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, *41*, 372-396.

252. Melton, A. W. The end-spurt in memorization curves as an artifact of the averaging of individual curves. *Psychol. Monogr.*, 1936, *47*, No. 212.

253. Melton, A. W. The methodology of experimental studies of human learning and retention: I. The functions of a methodology and the available

criteria for evaluating different experimental methods. *Psychol. Bull.* 1936, *33*, 305-394.

254. MELTON, A. W. Article on learning. In *Encyclopedia of educational research.* Monroe, W. S., Ed. New York: Macmillan, 1941.

255. MELTON, A. W., and IRWIN, J. MCQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, *53*, 173-203.

256. MELTON, A. W., and STONE, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, *30*, 295-310.

257. MELTON, A. W., and VON LACKUM, W. J. Retroactive and proactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, *54*, 157-173.

258. MILLER, J. The rate of conditioning of human subjects to single and multiple stimuli. *J. gen. Psychol.*, 1939, *20*, 399-408.

259. MILLER, N. E. Experimental studies of conflict. In *Personality and the behavior disorders.* Vol. I. Hunt, J. McV., Ed. New York: Ronald Press, 1944.

260. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, *38*, 89-101.

261. MILLER, N. E., and BUGELSKI, R. Minor studies of aggression: II. The influence of frustrations imposed by the in-group on attitudes expressed toward out-groups. *J. Psychol.*, 1948, *25*, 437-442.

262. MILLER, N. E., and DOLLARD, J. *Social learning and imitation.* New Haven: Yale University Press, 1941.

263. MINAMI, H., and DALLENBACH, K. M. The effect of activity upon learning and retention in the cockroach. *Amer. J. Psychol.*, 1946, *59*, 1-158.

264. MOORE, H. T. The genetic aspect of consonance and dissonance. *Psychol. Monogr.*, 1914, *17*, No. 73.

265. MORENO, J. L. Who shall survive? A new approach to the problem of human inter-relations. *Nervous and Mental Disease Monograph Series No. 58.* Washington, D. C.: Nervous and Mental Disease Publishing, 1934.

266. MORGAN, C. D., and MURRAY, H. A. A method for investigating fantasies: the Thematic Apperception Test. *Arch. Neurol. Psychiat., Chicago*, 1935, *34*, 289-306.

267. MORGAN, C. T. The hoarding instinct. *Psychol. Rev.*, 1947, *54*, 335-341.

268. MORGAN, C. T., and WALDMAN, H. "Conflict" and audogenic seizures. *J. comp. Psychol.*, 1941, *31*, 1-11.

269. MORGAN, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol., N. Y.*, 1916, *5*, No. 35.

270. MORGAN, J. J. B. Effect of non-rational factors on inductive reasoning. *J. exp. Psychol.*, 1944, *34*, 159-168.

271. MORTON, J. T. The distortion of syllogistic reasoning produced by personal convictions Ph.D. dissertation, Northwestern University, 1942.

272. MOWRER, O. H. A stimulus-response analysis of anxiety and its rôle as a reinforcing agent. *Psychol. Rev.*, 1939, *46*, 553-565.

273. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, *27*, 497-516.

274. Mowrer, O. H. An experimental analogue of "regression" with incidental observations on "reaction-formation." *J. abnorm. soc. Psychol.*, 1940, *35*, 56-87.
275. Mowrer, O. H., and Jones, Helen M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, *33*, 369-386.
276. Mowrer, O. H., and Ullman, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, *52*, 61-90.
277. Muenzinger, K. F. Motivation in learning: I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, *17*, 267-277.
278. Muenzinger, K. F., Bernstone, A. H., and Richards, L. Motivation in learning: VIII. Equivalent amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, *26*, 177-186.
279. Mukerji, N. P. An investigation of ability in work in groups and in isolation. *Brit. J. Psychol.*, 1940, *30*, 352-356.

280. Neifeld, M. R., and Poffenberger, A. T. A mathematical analysis of work curves. *J. gen. Psychol.*, 1928, *1*, 448-458.
281. Norris, Eugenia B., and Grant, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, *61*, 37-49.

282. Patrick, J. R. Studies in rational behavior and emotional excitement: I. Rational behavior in human subjects. *J. comp. Psychol.*, 1934, *18*, 1-22.
283. Patrick, J. R. Studies in rational behavior and emotional excitement: II. The effect of emotional excitement on rational behavior in human subjects. *J. comp. Psychol.*, 1934, *18*, 153-195.
284. Pavlov, I. P. *Conditioned reflexes.* Trans. by Anrep, G. V. London: Oxford University Press, 1927.
285. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, *30*, 93-113.
286. Perin C. T. A quantitative investigation of the delay-of-reinforcement gradient. *J. exp. Psychol.*, 1943, *32*, 37-51.
287. Perkins, C. C., Jr. The relation of secondary reward to gradients of reinforcement. *J. exp. Psychol.*, 1947, *37*, 377-392.
288. Peters, H. N. Experimental studies of the judgmental theory of feeling: III. The absolute shift in affective value conditioned by learned reactions. *J. exp. Psychol.*, 1939, *24*, 73-85.
289. Peters, H. N. The experimental study of aesthetic judgments. *Psychol. Bull.*, 1942, *39*, 273-305.
290. Peterson, J. Experiments in rational learning. *Psychol., Rev.* 1918, *25*, 443-467.
291. Philip, B. R., and Peixotto, Helen E. Generalization in the initial stages of learning nonsense syllables: I. Integral responses. *J. exp. Psychol.*, 1943, *33*, 50-63.
292. Plotkin, L. Stimulus generalization in Morse code. *Arch. Psychol., N. Y.*, 1943, *40*, No. 287.

293. Poffenberger, A. T. The effects of continuous mental work. *Amer. J. Psychol.*, 1927, *39*, 283-296.

294. Poffenberger, A. T. The effects of continuous mental work upon output and feelings. *J. appl. Psychol.*, 1928, *12*, 459-467.

295. Porter, E. H., and Biel, W. C. Alleged regressive behavior in a two-unit maze. *J. comp. Psychol.*, 1943, *35*, 187-195.

296. Porter, J. M., Jr. Backward conditioning of the eyelid response. *J. exp. Psychol.*, 1938, *23*, 403-410.

297. Postman, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.*, 1947, *60*, 101-108.

298. Postman, L., and Senders, Virginia L. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, *36*, 153-165.

299. Preston, M. G., Brotemarkle, R. G., and Campbell, E. G. Effect of change in motivation upon homogeneity of ergograms. *J. exp. Psychol.*, 1942, *31*, 497-504.

300. Proshansky, H., and Murphy, G. The effects of reward and punishment on perception. *J. Psychol.*, 1942, *13*, 295-305.

301. Razran, G. H. S. Studies in configural conditioning: VI. Comparative extinction and forgetting of pattern and of single-stimulus conditioning. *J. exp. Psychol.*, 1939, *24*, 432-438.

302. Reed, H. B. Fatigue and work curve from a 10-hour day in addition. *J. educ. Psychol.*, 1924, *15*, 389-392.

303. Reed, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, *36*, 71-87.

304. Reed, H. B. The learning and retention of concepts: IV. The influence of the complexity of the stimuli. *J. exp. Psychol.*, 1946, *36*, 252-261.

305. Reed, J. D. Spontaneous activity of animals. *Psychol. Bull.*, 1947, *44*, 393-412.

306. Reese, T. M. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.*, 1943, *55*, No. 251.

307. Renshaw, S., and Postle, Dorothy, K. Pursuit learning under three types of instruction. *J. gen. Psychol.*, 1928, *1*, 360-367.

308. Rethlingshafer, Dorothy. Effect of various concomitant activities upon the maintenance of a set. *J. exp. Psychol.*, 1945, *35*, 312-320.

309. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, *35*, 15-30.

310. Rice, P. B., Mowrer, O. H., and Allport, G. W. Symposium: the ego and the law of effect. *Psychol. Rev.*, 1946, *53*, 307-447.

311. Richter, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.*, 1922, *1*, No. 2.

312. Riker, B. L. The ability to judge pitch. *J. exp. Psychol.*, 1946, *36*, 331-346.

313. Robinson, E. S. Work of the integrated organism. In *A handbook of general experimental psychology.* Murchison, C. A., Ed. Worcester: Clark University Press, 1934.

314. Robinson, E. S., and Bills, A. G. Two factors in the work decrement. *J. exp. Psychol.*, 1926, *9*, 415-443.

315. Robinson, Mary F. The work decrement as affected by three kinds of meaningfulness. *J. exp. Psychol.*, 1938, *22*, 125-149.

316. Rodnick, E. H. Characteristics of delayed and trace conditioned responses. *J. exp. Psychol.*, 1937, *20*, 409-425.

317. Rodnick, E. H. Does the interval of delay of conditioned responses possess inhibitory properties? *J. exp. Psychol.*, 1937, *20*, 507-527.

318. Rodnick, E. H., and Klebanoff, S. G. Projective reactions to induced frustration as a measure of social adjustment. *Psychol. Bull.*, 1942, *39*, 489.

319. Roethlesberger, F. J., and Dickson, W. J. *Management and the worker.* Cambridge, Mass: Harvard University Press, 1940.

320. Rohrer, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, *37*, 473-493.

321. Rosenthal, B. G. Hypnotic recall of material learned under anxiety- and non-anxiety-producing conditions. *J. exp. Psychol.*, 1944, *34*, 369-389.

322. Rosenzweig, M. Discrimination of auditory intensities in the cat. *Amer. J. Psychol.*, 1946, *59*, 127-136.

323. Rosenzweig, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, *32*, 64-74.

324. Rothe, H. F. Output rates among butter wrappers: I. Work curves and their stability. *J. appl. Psychol.*, 1946, *30*, 199-211.

325. Rotter, J. B. Level of aspiration as a method of studying personality: I. A critical review of methodology. *Psychol. Rev.*, 1942, *49*, 463-474.

326. Ryan, A. H., and Warner, Mary. The effect of automobile driving on the reactions of the driver. *Amer. J. Psychol.*, 1936, *48*, 403-421.

327. Ryan, T. A. *Work and effort.* New York: Ronald Press, 1947.

328. Saloman, Ann D. Visual field factors in the perception of direction. *Amer. J. Psychol.*, 1947, *60*, 68-88.

329. Sandahl, Robert. The effect of distribution of learning on lists of synonymous adjectives. M.A. thesis, Northwestern University, 1948.

330. Sanders, Marjorie J. An experimental demonstration of regression in the rat. *J. exp. Psychol.*, 1937, *21*, 493-510.

331. Sanford, R. N. The effects of abstinence from food upon imaginal processes: a preliminary experiment. *J. Psychol.*, 1936, *2*, 129-136.

332. Sanford, R. N. The effects of abstinence from food upon imaginal processes: a further experiment. *J. Psychol.*, 1937, *3*, 145-159.

333. Schieber, Ruth. Retention as a function of stage of practice. Unpublished.

334. Schlosberg, H. Conditioned responses in the white rat. *J. genet. Psychol.*, 1934, *45*, 303-335.

335. Schmidt, H. O. The effects of praise and blame as incentives to learning. *Psychol. Monogr.*, 1941, *53*, No. 240.

336. Sears, R. R. Initiation of the repression sequence by experienced failure. *J. exp. Psychol.*, 1937, *20*, 570-580.

337. Sears, R. R. Success and failure: a study of motility. In *Studies in personality.* McNemar, Q., and Merrill, Maud, Eds. New York: McGraw-Hill, 1942.

338. SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Com. Bull.*, 1943, No. 51.

339. SEARS, R. R. Experimental analysis of psychoanalytic phenomena. In *Personality and the behavior disorders.* Vol. I. Hunt, J. McV., Ed. New York: Ronald Press, 1944.

340. SEARS, R. R., and HOVLAND, C. I. Experiments on motor conflicts: II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, *28*, 280-286.

341. SEARS, R. R., HOVLAND, C. I., and MILLER, N. E. Minor studies of aggression: I. Measurement of aggressive behavior. *J. Psychol.*, 1940, *9*, 275-295.

342. SEASHORE, C. E., LEWIS, D., and SAETVEIT, J. G. Revision of the Seashore Measures of Musical Talents. *Univ. Ia. Stud. Aims Progr. Res.*, 1940, *65*, No. 388.

343. SEASHORE, R. H. Work methods: an often neglected factor underlying individual differences. *Psychol. Rev.*, 1939, *46*, 123-141.

344. SEASHORE, R. H., UNDERWOOD, B. J., HOUSTON, R., and BERKS, L. The influence of knowledge of results on performance. Unpublished.

345. SELLS, S. B. The atmosphere effect: an experimental study of reasoning. *Arch. Psychol., N. Y.*, 1936, *28*, No. 200.

346. SEWARD, J. P. An experimental study of Guthrie's theory of reinforcement. *J. exp. Psychol.*, 1942, *30*, 247-256.

347. SEWARD, J. P., DILL, Jane B., and HOLLAND, Mildred A. Guthrie's theory of learning: a second experiment. *J. exp. Psychol.*, 1944, *34*, 227-238.

348. SHAFFER, L. F. *The psychology of adjustment.* New York: Houghton Mifflin, 1936.

349. SHARP, Agnes A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, *22*, 395-418.

350. SHAW, W. A., NEWMAN, E. B., and HIRSH, I. J. The difference between monaural and binaural thresholds. *J. exp. Psychol.*, 1947, *37*, 229-242.

351. SHERIF, M., and CANTRIL, H. *The psychology of ego-involvements.* New York: Wiley, 1947.

352. SHERMAN, M., and JOST, H. Frustration reactions of normal and neurotic persons. *J. Psychol.*, 1942, *13*, 3-19.

353. SHIPLEY, W. C., COFFIN, Judith I., and HADSELL, Kathryn C. Affective distance and other factors determining reaction time in judgments of color preference. *J. exp. Psychol.*, 1945, *35*, 206-215.

354. SHIPLEY, W. C., NORRIS, Elizabeth D., and ROBERTS, Margaret L. The effect of changed polarity of set on decision time of affective judgments. *J. exp. Psychol.*, 1946, *36*, 237-243.

355. SIIPOLA, Elsa M. Implicit or partial reversion-errors: a technique of measurement and its relation to other measures of transfer. *J. exp. Psychol.*, 1940, *26*, 53-73.

356. SKINNER, B. F. *The behavior of organisms; an experimental analysis.* New York: D. Appleton-Century, 1938.

357. SMITH, G. H. Attitudes toward Soviet Russia: I. The standardization of a scale and some distributions of scores. *J. soc. Psychol.*, 1946, *23*, 3-16.

358. SMITH, G. H. Attitudes toward Soviet Russia: II. Beliefs, values, and other

characteristics of pro-Russian and anti-Russian groups. *J. soc. Psychol.*, 1946, *23*, 17-33.

359. SMITH, G. H. Beliefs in statement labelled fact and rumor. *J. abnorm. soc. Psychol.*, 1947, *42*, 80-90.

360. SOLOMON, R. L. Effort and extinction rate: a confirmation. *J. comp. physiol. Psychol.*, 1948, *41*, 93-101.

361. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, *45*, 1-40.

362. SPENCE, K. W. Theoretical interpretations of learning. In *Comparative psychology*. Moss, F. A., Ed. New York: Prentice-Hall, 1942.

363. SPENCE, K. W. The rôle of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, *54*, 1-8.

364. SPENCE, K. W., and LIPPITT, R. An experimental test of the signgestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, *36*, 491-502.

365. SPOONER, Alice, and KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, *60*, 321-334.

366. SPROW, A. J. Reactively homogeneous compound trial-and-error learning with distributed trials and terminal reinforcement. *J. exp. Psychol.*, 1947, *37*, 197-213.

367. STEVENS, S. S., and VOLKMANN, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.*, 1940, *53*, 329-353.

368. STONE, G. R. The effect of negative incentives in serial learning: III. Fixation due to an isolated verbal punishment. *J. gen. Psychol.*, 1948, *38*, 207-216.

369. STROUD, J. B. The rôle of muscular tensions in stylus maze learning. *J. exp. Psychol.*, 1931, *14*, 606-631.

370. STROUD, J. B. Experiments on learning in school situations. *Psychol. Bull.*, 1940, *37*, 777-807.

371. STROUD, J. B. The rôle of practice in learning. *Forty-First Yearbook, Nat. Soc. Study Educ.*, Part II. Bloomington, Ill.: Public School Publishing, 1942.

372. SWITZER, S. A. Backward conditioning of the lid reflex. *J. exp. Psychol.*, 1930, *13*, 76-97.

373. SWITZER, S. A. Disinhibition of the conditioned galvanic skin response. *J. gen. Psychol.*, 1933, *9*, 77-100.

374. SWITZER, S. A. Anticipatory and inhibitory characteristics of delayed conditioned reactions. *J. exp. Psychol.*, 1934, *17*, 603-620.

375. TELFORD, C. W., and STORLIE, A. The relation of respiration and reflex winking rates to muscular tension during motor learning. *J. exp. Psychol.*, 1946, *36*, 512-517.

376. TERMAN, L. M., and MERRILL, Maud A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.

377. THORNDIKE, E. L. Mental fatigue. *J. educ. Psychol.*, 1911, *2*, 61-80.

378. THORNDIKE, E. L. *The fundamentals of learning.* New York: Bureau of Publications, Teachers College, Columbia University, 1932.

379. THORNDIKE, E. L. An experimental study of rewards. *Teach. Coll. Contr. Edu.*, 1933, No. 580.

380. THORNDIKE, E. L. *The psychology of wants, interests, and attitudes.* New York: D. Appleton-Century, 1935.

381. Thornton, G. R. The effect upon judgments of personality traits of varying a single factor in a photograph. *J. soc. Psychol.*, 1943, *18*, 127-148.

382. Thornton, G. R. The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *J. appl. Psychol.*, 1944, *28*, 203-207.

383. Thune, L. E., and Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, *32*, 185-200.

384. Thurstone, L. L. Psychophysical methods. In *Methods of psychology*. Andrews, T. G., Ed. New York: Wiley, 1948.

385. Thurstone, L. L., and Chave, E. J. *The measurement of attitude; a psychophysical method and some experiments with a scale for measuring attitude toward the church.* Chicago: University of Chicago Press, 1929.

386. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

387. Tolman, E. C., Ritchie, B. F., and Kalish, D. Studies in spatial learning: I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, *36*, 13-24.

388. Tolman, E. C., Ritchie, B. F., and Kalish, D. Studies in spatial learning: II. Place learning versus response learning. *J. exp. Psychol.*, 1946, *36*, 221-229.

389. Trowbridge, Margery H., and Cason, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, *7*, 245-260.

390. Underwood, B. J. Associative inhibition in the learning of successive paired-associate lists. *J. exp. Psychol.*, 1944, *34*, 127-135.

391. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, *59*, No. 273.

392. Underwood, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, *38*, 29-38.

393. Underwood, B. J. "Spontaneous recovery" of verbal associations. *J. exp. Psychol.*, 1948, *38*, 429-439.

394. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, in press.

395. Walker, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, *38*, 39-49.

396. Wallach, H., and Henle, Mary. An experimental analysis of the law of effect. *J. exp. Psychol.*, 1941, *28*, 340-349.

397. Wallen, R. Ego-involvement as a determinant of selective forgetting. *J. abnorm. soc. Psychol.*, 1942, *37*, 20-39.

398. Wang, G. H. The relation between "spontaneous" activity and oestrous cycle in the white rat. *Comp. Psychol. Monogr.*, 1923, *2*, No. 6.

399. Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, *49*, No. 220.

400. Weber, Henrietta, and Wendt, G. R. Conditioning of eyelid closure with various conditions of reinforcement. *J. exp. Psychol.*, 1942, *30*, 114-124.

401. Wedell, C. H. The nature of the absolute judgment of pitch. *J. exp. Psychol.*, 1934, *17*, 485-503.

402. Weitz, J. A further study of the relation between skin temperature and cutaneous sensitivity. *J. exp. Psychol.*, 1942, *30*, 426-431.

403. Welch, L. The transition from simple to complex forms of learning. *J. genet. Psychol.*, 1947, *71*, 223-251.

404. Wertheimer, M. *Productive thinking.* New York: Harper, 1945.

405. White, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, *50*, 157-186.

406. Whiteley, P. L., and McGeoch, J. A. The curve of retention for poetry. *J. educ. Psychol.*, 1928, *19*, 471-479.

407. Wickens, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *J. exp. Psychol.*, 1938, *22*, 101-123.

408. Wickens, D. D. Studies of response generalization in conditioning: I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, *33*, 221-227.

409. Wickens, D. D., and Wickens, Carol D. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, *26*, 94-102.

410. Wickens, D. D., and Wickens, Carol D. Some factors related to pseudo-conditioning. *J. exp. Psychol.*, 1942, *31*, 518-526.

411. Wilkins, Minna C. The effect of changed material on ability to do formal syllogistic reasoning. *Arch. Psychol., N. Y.*, 1928, *16*, No. 102.

412. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, *11*, 21-29.

413. Williams, M. The effects of experimentally induced needs upon retention. Ph.D. dissertation, Northwestern University, 1948.

414. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, *23*, 506-522.

415. Wilson, J. T. The formation and retention of remote associations in rote learning. Ph.D. dissertation, Stanford University, 1948.

416. Withey, S. B. Relationship between reminiscence effects and rest interval activity in serial anticipation learning. M.A. thesis, Northwestern University, 1947.

417. Witmer, Louise R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, *47*, 337-359.

418. Wolfe, J. B. Effectiveness of token-rewards for chimpanzees. *Comp. Psychol. Monogr.*, 1936, *12*, No. 60.

419. Wolfle, Helen M. Time factors in conditioning finger-withdrawal. *J. gen. Psychol.*, 1930, *4*, 372-378.

420. Wolfle, Helen M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, *7*, 80-103.

421. Woodward, Patricia. An experimental study of transfer of training in motor learning. *J. appl. Psychol.*, 1943, *27*, 12-32.

422. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

423. Wyatt, Ruth F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, *58*, No. 267.

424. YOUNG, P. T. *Motivation of behavior.* New York: Wiley, 1936.

425. YOUNG, P. T. Studies of food preferences, appetite and dietary habits: VII. Palatability in relation to learning and performance. *J. comp. physiol. Psychol.*, 1947, *40*, 37-72.

426. ZANDER, A. F. A study of experimental frustration. *Psychol. Monogr.*, 1944, *56*, No. 256.

427. ZIRKLE, G. A. Success and failure in serial learning: II. Isolation and the Thorndike effect. *J. exp. Psychol.*, 1946, *36*, 302-315.

Author Index

If author's name does not appear on page in text where reference is cited, the reference number appears in parentheses here after the page number.

Glossary and Subject Index

Absolute zero, 83, 87

Action Potentials — *Electrical potentials generated in the muscles by contraction of the fibers:* 165

Active recitation, influence on learning, 418-419

Additivity — *A property of extensive scales which may be operationally exemplified by adding two equal units together by usual arithmetic operations to obtain a resultant quantity which is twice as great as either alone:* 83

Adjustment, method of, *see* Average Error, Method Of

Aesthesiometer — *Device used to stimulate the skin to determine the two-point threshold of tactual discrimination:* 58

Affective judgments, experiments on, 97, 107-113, 121-123

Affectivity, influence on multiple-response learning, 412-414; on retention, 531-536; steps in experimentation, 412-414

Aggression — *Response to frustration which, in its most restricted sense, refers to tendency for S to attack source of frustration:* 207-208, 211, 213, 214, 216, 232; *see also* Displaced Aggression

Ambiguity, in constructing attitude scale, 91-93

Ambivalence — *Opposed response tendencies toward a single person or object:* 109, 244

Amount of material, influence on learning, 419

Anchoring Effect — *Subjective standard providing a "value" against which other stimuli are judged:* 110-113, 140

Anxiety — *Motive produced by fear of punishment:* 192, 278

Apathy — *Response to frustration in which motivation deteriorates; a form of psychological withdrawal:* 208, 224

Apparent Movement — *Perceived movement which objectively does not take place:* 42

Approach-Approach Conflict — *Stimuli for two or more incompatible approach responses are presented simultaneously:* 243 f.

methods for inducing, 247-248

Approach-Avoidance Conflict — *Stimuli for an approach and an avoidance response are presented simultaneously:* 243 f.

methods for inducing, 248-250

Approach Gradient — *Description of the observation that the nearer an organism is to a positive incentive the greater the "pulling power" of that incentive:* 272-275

Approach Response — *Response toward a positive incentive:* 242

Areas of study, integration of, 17-21

Assembly tasks, in work, 559

Assimilation — *Process inferred to be operating when sensory judgments, e.g., judgments of visual extents, are positively influenced by characteristics of other stimuli in the environment:* 75

Atmosphere Effect — *Unanalyzed bias operating in syllogistic reasoning which causes S to accept more readily a conclusion stated in the "atmosphere" (positive or negative) of the premises than when not so stated:* 455-457

Scale, Extensive — *Dimension which has been quantified so that all units are equal, which has an absolute zero, and additive properties:* 81-84
Scale, Intensive — *Dimension which has been quantified only in terms of rank order of magnitude of characteristic under consideration:* 81, 84
Scientific attitude, description of, 21
Scientific Generalization — *Extension of empirical findings beyond specific conditions and subjects from which findings were derived:* 337-339
Search, in thinking, 430
Seashore Measures of Musical Talent, 80, 119
Secondary Reinforcement — *Receipt of symbolic incentive:* 171-176, 359, 360, 367, 492, 505-506, 526-527
Secondary Variables in Learning — *Variables which cause rate of learning to vary but absence of which does not prohibit learning from taking place:* 467 f.
Sense modality, influence on learning, 419, 427
Sensory pre-conditioning, 364, 506
Sentence completion tests, 560
Serial Learning — *Learning in which A is the stimulus for response B, B the stimulus for response C, C the stimulus for D, etc.:* 294, 386-387, 419-425
Serial position, 296
Set — (1) *Relationship between formal instructions given* S *and performance following instructions; or* (2) *Relationship between previously acquired methods of solving problems and current performance in solving problems:* 189-190, 447
  experiments on variation of, 189-192, 447-455
  influence on thinking, 447-455; on work, 574-577
  polarity of, 99
Sex motive, 163
Short-Trace Conditioning — *Conditioned stimulus precedes but does not overlap unconditioned stimulus and time between cessation of conditioned stimulus and onset of unconditioned stimulus is one-tenth to five seconds:* 353, 355
Similarity — *Degree of homogeneity among objects or symbols; must be defined operationally by dimensionalizing the many attributes within which degrees of homogeneity may prevail:* 286 f.
  and reminiscence, 521
  as defined by formal identity, 293-301; by meaningfulness, 301-302
  dimensional variations in, 292-304
  formal, in motor task, 296, 586; in figural designs, 298-299; in paired nonsense syllables, 293; in serial verbal tasks, 294-295
  in code learning, 409
  in perceptual learning theory, 477-478
  in transfer of work decrement, 586-587
  influence on multiple-response learning, 407-410; on retention, 530-531; on retroactive inhibition, 546-547; on thinking, 445-447; on transfer, 286-314; on work, 580-581
  non-dimensional variations in, 304-313
Simple T-Maze — *Maze, used with animals, constructed like a T, thus having a single choice point:* 168, 250, 265, 312, 349, 360, 367, 476, 492, 498
Simultaneous Conditioning — *Conditioned stimulus precedes and overlaps unconditioned stimulus and time between onset of conditioned stimulus and unconditioned stimulus may vary from zero to five seconds:* 353-355
Single Stimuli, Method Of — *Method for studying discriminal processes by presenting* S *with successive stimulus and requesting a judgment concerning these stimuli along a specified dimension;* sometimes called *method of absolute judgment:* 71, 106 f., 122, 124
  illustrative experiments, 107 f.
  summary, 114
Skin color, experiment on judgment of, 138-140
Skinner Box — *Device used to study problems of learning and motivation with*